



THE UNIVERSITY *of* EDINBURGH

Title	Epistemic constraints on semantic theory
Author	Smith, Barry Crawford
Qualification	PhD
Year	1991

Thesis scanned from best copy available: may contain faint or blurred text, and/or cropped or missing pages.

Digitisation notes:

- p.154 missing from original numeration.

EPISTEMIC CONSTRAINTS ON SEMANTIC THEORY

Barry Crawford Smith

**DOCTOR OF PHILOSOPHY
UNIVERSITY OF EDINBURGH**

1989



I declare that the work I have presented here is my own.

Barry Crawford Smith

Acknowledgements

Thanks are best said privately, and they will be. I have many, many people to thank. But I should just like to address words of gratitude to the following people in public. Firstly, to Neil Tennant who provided the inspiration and constant encouragement to engage with the issues in the most philosophically exacting way. His tireless and infectious enthusiasm has proved an enduring spur to greater philosophical effort. Thanks are due also to the whole Tennant family, Meg, Lizzie, Kate and Neil for the much needed and much appreciated warmth of Landrick. I should also like to thank Gill Brown for giving me a lasting and firm appreciation of empirical approaches to language; and I should also like to thank her for her support and conviction in the worth of my developing ideas. I should like to thank David Hamlyn for gently given good advice and strong professional support.

Thanks are due also Lisa Robinson and Jeremy Shaw for the enormous efforts they have put in on my behalf.

In the early days of postgraduate study I was struck by the clarity and depth of engagement in the writings of Crispin Wright. Subsequently, I had the opportunity to discuss philosophy with him and was pleased to find a counter-example to the 'reading them' and 'meeting them' wisdom in philosophy. I have enjoyed all of those quick and keen discussions, and learned a lot from his generosity of thought and gladly given insight. He sets an important example of how philosophy should be done. My thanks for those times and efforts. Thanks go also to Geoffrey Madell and Dore Scaltsas for their patience and perseverance.

My greatest thanks are reserved for the following two people. My thanks and love to Deborah Seyman, without whom not. Thanks lastly, but in great measure, to my colleague Martin Davies whose keen good sense and extraordinary philosophical talent were given out unstintingly over the period of writing up. I have benefited hugely from his professional, intellectual and moral support; with admiration, gratitude and friendship, my heartfelt thanks.

ABSTRACT

The thesis examines a certain philosophical assumption that meanings and mental states have to be attributed to speakers conjointly in attempting to gain an understanding of their behaviour, linguistic and otherwise. It is shown that this attributive view of mind and meaning cannot do justice to the full range and character of our knowledge of language. Linguistic knowledge is a property of individuals and any account of it must draw upon a psychological theory of their linguistic abilities. The possibility of a descriptive theory of meaning eschewing all mention of cognitive psychological facts is considered and rejected. The conclusion is reached that theorising about language that accommodates speakers' knowledge goes well beyond the brief of the a priori theorist's project of interpretation or description and points to the need for a more substantive treatment of the nature of mental states.

Table of Contents

CHAPTER ONE: A PHILOSOPHY OF MIND AND LANGUAGE

1.1.	A Priori and Empirical Conceptions of Psychology	1
1.2	Psychological Attributionism	6
1.3	The Three Stances	9
1.4	The Involvement of the Mind in Speech: A Prior Assumption	11
1.5	Setting Aside Grice	12
1.6.	Setting Aside Model-Theoretic Semantics	13
1.7	Interpretation as a Philosophy of Mind and Language?	18
1.8	Advocates of Interpretation	19

CHAPTER TWO: INTERPRETATION AND THE MENTAL

2.1	Introduction	20
2.2	The Strategy of Interpretation	20
2.3	The Principles of Interpretation	23
2.4	The Holism of the Mental	28
2.5	A Third-Person Perspective	31
2.6	The Narrative Character of Interpretation	32
2.7	Rational and Causal Explanation	41
2.8	The Irreducibility of the Mental	47

CHAPTER THREE: INDETERMINACY AND THE MENTAL

3.1	Introduction	60
3.2.	Irreducibility to Indeterminacy	61
3.3	Indeterminacy Between Interpretational Schemes	70
3.4	Indeterminacy Within an Interpretational Scheme	72

CHAPTER FOUR: INTERPRETATION AND MEANING

4.1	The Theory of Meaning: a Fregean Overview	84
4.2	Truth and Interpretation	87
4.3	Empirical Constraints on Truth Theory	90
4.4	Psychological Constraints on Truth Theory	98
4.5	Putting the Empirical Constraints to Work	107

Table of Contents

4.6	Formal Constraints on Truth Theory	111
4.7	Truth Theory and the Physical Facts	115
4.8	Systematicity and Epistemological Holism	118
CHAPTER FIVE: INDETERMINACY AND MEANING		129
5.1	Meaning and Logical Form	129
5.2	Why Should Davidson's Semantic Theory be Structure-Discerning?	134
5.3	The Inscrutability of Reference	143
5.4	Logical Form as Significant Semantic Structure?	159
5.5	T-Theories, Sentence Meanings and Individual Beliefs	180
CHAPTER SIX: PHILOSOPHY, MIND AND LANGUAGE		
6.1	Theories of Meaning and Theories of Understanding	198
6.2	How Should We Construe the Theory of Meaning?	205
6.3	Not a Theory of Meaning, not a Theory of Understanding	219
6.4	Linguistic Abilities as Cognitive Psychological Abilities	230
CHAPTER SEVEN: THE INDISPENSIBILITY OF CONCEPTS		
7.1	The System of Concepts	246
7.2	Causal Systematicity	250
CHAPTER EIGHT: KNOWLEDGE OF LANGUAGE		
8.1	Knowledge of Language and Theories of Language	263
8.2	Knowledge of Syntax	272
8.3	Principles and Parameters	283
8.4	Government/ Binding Theory	284
8.5	Syntax and the Three Stances	292
CONCLUSIONS		324

Chapter One

A Philosophy of Mind and Language

1.1) A Priori and Empirical Conceptions of Psychology

I begin with a prevalent philosophical conception of mind and language, which could be called, rather broadly, a Davidsonian view. This is the view that the study of mind and meaning should be reserved for philosophy alone. While it will be conceded that the empirical facts studied in the cognitive and brain sciences make possible and sustain our intentional phenomena, it will be the claim of this conception that no investigation of that empirical domain can cast light on the form and character of minds and meanings. Theories of mind and theories of meaning are philosophical theories, unilluminated by scientific explanations. The beliefs, desires and meanings they study are fully characterised and also constituted by the philosophical theories that describe them. This is the view I mean to challenge.

The a priori claim will be that because we ordinarily know the minds of others, we ordinarily know what minds are. Philosophical reflection will show that our ordinary reasons for supposing ourselves to have knowledge of other people's minds provides our only access to an understanding of the mental. Proponents of such a view will suppose that we attribute minds to others when we judge them to be agents and assess their behaviour in rational terms. The principles of assessment must be the a priori principles by which we interpret others, and so the concept of mind will be exhaustively characterised by the formal and empirical conditions on our methods of interpretation. I shall call someone who thinks like this an attributionist; where an attributionist is one who thinks the beliefs and desires we attribute to rational agents are not the objects of the interpreter's inquiry, but the products of it.

An attributionist about the mind must also be an attributionist about meaning, given the intimate link between minds and meanings. Meanings, too, will arise in the context of interpretation; in particular, in

Chapter One

that part of interpretation that addresses people's linguistic behaviour. The meanings of their words, like their states of belief and desire, will be a matter of attribution. They result from the precise application of the semantical and psychological categories we have devised to classify and evaluate people's behaviour.

There is a certain modesty in this stance: it presumes that our ordinary ways of thinking are good enough to tell us all there is to know about minds and meanings. However, the philosophical theory that reflects on these "ordinary ways" leads to a flagrantly metaphysical picture of minds and a degenerate conception of meaning as a result of its commitments to a number of counter-intuitive consequences. It is one thing to give prominence to the notion of commonsense in philosophy, but quite another to propose a philosophical conception of it. And it is with respect to the second of these strands of attributionism that we must test the doctrine.

I shall argue that there is something altogether faulty in such a conception, and that its puzzling consequences result from trying to stick too narrowly to commonsense to tell us what sorts of creatures we are and what linguistic and psychological resources are available to us. We need a less protectionist account of the complex phenomena of mind and language that give rise to, and give form to, the visible properties of speech and thought. Commonsense itself tells us that not everything we want to know about the world and our place in it can be discovered from the armchair. I shall suggest that the concept of mind and the concept of language are like the concept *gold* in having both an everyday and a more technical use, without being ambiguous. A full understanding of the folk notion of a language, say, is not taken to require knowledge of the more technical use made of it by a linguist; and yet if we want to know in precise terms what it is for something to be a language, or to be this or that particular language we should turn to the experts. The same would go for our use of the concept of mind, although in that case, as was once true of 'gold', we are still in the process of developing a more theoretically precise understanding of the concept. This is the phenomenon Hilary Putnam calls 'the division of linguistic labour'. It neither dispenses with our

ordinary, folk understanding of a concept (or expression), nor denies the importance of theoretical refinements to it. This is the model I have of the relation of folk psychological notions to those of linguistic and cognitive science. The notions of mind and language are open to both ordinary reflection and scientific inquiry. What is needed, therefore, is a way of combining philosophical reflection with empirical investigation of our distinctive linguistic and cognitive capacities; a way which does not lead us into philosophical error. We need a conception of mind and meaning that is consistent with a more substantive account of psychological explanation; this will mean accommodating semantics and psychology in an explanatory framework where intelligent behaviour is explained by reference to cognitive facts which are at once both genuinely mental but beyond the ordinary interpreter's reach. Thus it will not be by deploying a cognitive theory that the interpreter is able to know what he does of other people's minds and meanings, but rather it will be by his satisfying such a theory. He will not have to know of the presence of these complex cognitive states, he will have to be in them. In this way, I shall prise apart the roles of commonsense reasoner and psychological theorist, denying the purported a priori link between the ordinary epistemology of minds and explanation of the mental.

The case I shall focus on is our knowledge of language and its role in explanations of linguistic behaviour. I shall draw upon recent work by Noam Chomsky that suggests that an account of our knowledge of language is the study of a complex cognitive structure; one module of the mind. Knowledge of language is a complex mental phenomenon which most of us possess. It requires detailed empirical study at more than one level of cognitive organisation; and for these reasons I shall argue that the nature and structure of that knowledge is not simply a matter of what we attribute to speakers, ordinarily, when we try to make sense of their talk; nor is it, less simply, a matter of rational reconstruction of their linguistic practice under the guidance of certain a priori theoretical principles. It is a substantive property of the human mind, not merely an attributed one; it is sustained by a complex organisation of cognitive states whose influence on our behaviour is owed to their being neither simply reasons nor causes

of that behaviour. The effect of these states on linguistic practice is owed to the nature of mental representations which participate in the mental processes that result in the production and comprehension of speech. Specific mental representations are invoked when we give content-using explanations in computational psychology. These are neither physical explanations of a speaker's output, nor the folk psychological explanations of the ordinary interpreter, and for this reason we must invoke an intervening level of cognitive psychological organisation to which the mentioned states and processes belong.

It is not my intention to claim that this intervening level must replace the ordinary intentional levels of description and explanation, rather the scientific cognitivist's claim is that facts about the underlying cognitive states and processes helps to explain and illuminate the form and character of the behaviour we can characterise in those intentional terms. The cognitivist will insist that a full explanation of the intentional phenomena must make reference to the cognitive psychological structures that subserve them.

The above claim is intended to steer one clear of a potential misunderstanding of my project. To claim that certain facts about a speaker's cognitive apparatus underpin his ability to use a language is quite consistent with claiming that the contents a language user can express are not entirely constituted by facts internal to the linguistic processor. There is no immediate threat here of a Cartesian retreat to solipsism. The contents of a person's speech and thought depend on his contact with a social and physical environment. But his ability to retrieve the significance of information he encounters there depends in part on his psychological make up. The full task of the project I want to make room for would have to explain how it is that creatures with the relevant cognitive apparatus make their minds and meanings available to one another in a public language. My contribution will be to show that neglect of the empirical study of the cognitive component leaves one with an account that can furnish us with very little of what people say or think at all.

Chapter One

Now, before we go any further, I think it is important to dispel one looming misconception: that this is no longer a philosophical question. I dispute this. I am not changing the subject, or reverting to a merely empirical issue. The question of which empirical details bear on the nature of mind and language is a philosophical question. As Donald Davidson himself admits when discussing which creatures have propositional attitudes:

The question is not empirical; the question is what sort of empirical evidence is relevant to deciding when a creature has propositional attitudes. (1982 p317)

I would add, that the question is also what sort of evidence is relevant to deciding *which* particular propositional attitudes a creature has. So long as we see things in this way, there is still room for an a priori, philosophical dispute about the range of empirical conditions on psychological ascription. This is still a philosophical dispute.

I shall argue that the underlying cognitive states and processes in individuals shape the cognitive products they produce, and that by gaining a clearer understanding of the precise nature of such processes, we shall be in a position to learn something philosophically important about why we may interpret someone's behaviour in the intentional terms that we do. To this extent, the underlying contingencies, and not just the philosophical descriptions imposed on them, give character and form to a person's intentional behaviour.

I shall seek to establish the grounds for a more substantive account of what it is to have a certain belief, to utter a sentence with a particular meaning, or to possess a certain concept. The position I shall take for the linguistic case is neatly expressed in the following remarks by the psychologist of language, Herbert Clark:

What speakers produce is necessarily constrained by the processes by which they produce what they produce, so theories of language products must ultimately be constrained by theories of language processes. To picture philosophy of language as immune from empirical evidence is therefore a distortion of

fact. The issue, rather, is what facts matter and how. (Clark 1986, p17)

Knowledge of language is the subject of study for philosophers, linguists and cognitive psychologists working together: its study is a proper task of the cognitive sciences.

1.2) Psychological Attributionism

In striking contrast, the attributionist's view of empirical science is unequivocal: the mental and the meaningful are constituted by the intentional categories we impose on a creature to make sense of its behaviour in rational terms. Lower levels of description reveal mere contingencies about the internal workings of a creature and cannot hope to explain the rational patterns of intentional ascription that bring the (rational) mind to light. Thus empirical investigations of underlying mechanisms, and the sub-personal states and processes that are causally responsible for behaviour, can tell us nothing about the mental categories for classifying it.

However, even on this a prioristic conception of the mental there must be some relation between the higher level of intentional description and the underlying empirical facts, as Davidson's remarks indicate. And it will be in apprehending this relation that the attributer decides, when confronted with the mere contingencies, whether to apply or withhold his mental terminology. So there will have to be some physical or behavioural constraints on psychological attributions, and for the attributionist, it is the project of interpretation that supplies all the necessary links: it must provide the a priori means both to characterise the intentional realm and to effect the transition from the underlying physical level to the level of intentional characterisation. The project describes an intentional structure bound by rational explanations of behaviour. Each single attribution depends on a background of others and must belong to an overall delineation of the thinker's psychological states; this total scheme provides a conception of a thinker with reference to which we can interpret his talk and explain his actions. An assumption of rationality is

Chapter One

built into this overall psychological conception. Rationality is conceived to be constitutive of the range of application of the concepts of the propositional attitudes, and so to a large extent the mental life of any individual subject must be rational. Rationality is a normative standard of the interpreter, an ideal he aims for in constructing an interpretational scheme.

The idea that subjects must be to a large extent rational may seem to be threatened by the idea that for the attributionist the propositional attitudes, and the meanings of individual sentences, are constructs of the interpreter, ascribed to people and their utterances to explain their verbal and non-verbal behaviour. The idealisation may seem to miss the mark. But it is in the nature of attributionist thinking to suppose that it is only when someone's behaviour can be brought under the terms of such a rational description that the person has any claim to mental life at all.

This is to see mental states as a matter of what intentional descriptions can be given to a person's behaviour from the interpreter's point of view. Thus mental states are forged from behaviour by interpreting it. And although this makes them depend essentially on facts about third-person interpretation, the mental states are given some life of their own by describing them as states in the agent that cause his actions. At first this may seem to contradict the attributionist's ideology. But strictly speaking, what happens here is that there are internal physical parts of a person that cause his physical movements, and it is when we can re-describe that behaviour as intentional action that we are given licence to re-describe those causally relevant inner states as beliefs and desires. Our right to do so is secured by the philosophical thesis that reasons are the causes of intentional behaviour: only in so far as we are entitled to describe people as acting intentionally are we entitled to describe the causes of their behaviour as mental states, and only in so far as behaviour is caused by mental states can we claim it is intentional action. This thesis gives us no prior route to the discovery of the cause or the nature of the effect: we must simply work as interpreters to earn the right to impose descriptions of intentional states and actions on physical causes of bodily movements. The causes of physical change will be there anyway, and their subsequent

intentional re-descriptions depend entirely on the possibilities of interpretation at the higher level; thus intentionality has nothing to do with intrinsic features of the states themselves. An event is only mental because it is so described. Nothing in a physical cause accounts for its intentional character; this remains in the eye of the interpreter. Nevertheless, the intentional states we ascribe are ultimately physical states: states of the world we describe in mental terms. The attributionist is still a monist.

The pure attributionist thesis comes down to this: inner mental life, as delineated by beliefs, desires, hopes and fears, is just that part of the physical world that is answerable to an intentional taxonomy imposed from the outside. The taxonomy belongs to a priori psychology: the psychology of ordinary belief-desire attribution. The principles of propositional attitude attribution are constitutive of this psychology, and so constitutive of the intentionally taxonomised mental states themselves.

I shall call this position *pure* attributionism, since a number of impure versions of this doctrine exist.

An attributionist is one who believes that in a theory of mental states, a theory of the external conditions for their attribution is prior to a theory of what it is to possess them; indeed, he believes that theories of the former kind provide the only route to the latter. But the pure attributionist would find the question of what it is to possess a mental state misconceived; at least misconceived if it was thought that there was something independent of interpretation *in virtue of which* the attribution to a subject of a given mental state was true. For the pure attributionist there can be no substantive answer to the question, 'What is it to possess a given belief (or concept)?' over and above being told, it is to be apt to be ascribed that state of belief (or that concept) in an interpretation that makes "the best total sense possible..... of the life and conduct" of the thinker (see Wiggins 1980, p199). On this understanding of the attributionist, he believes that mental states can only be what we attribute to a subject when we try to make sense of him in rational terms, and that we make sense of people in accordance with certain principles of interpretation considered constitutive of the mental. These principles are

a priori: we do not test them by how accurately they characterise the mind of another, instead our characterisations of people's mental life are tested against the principles. They give our rational and only grounds for making psychological ascriptions, and they exhaust the content of concepts used in those ascriptions. For the pure attributionist there can be no gap between the mind of a thinker and the best interpretation to be given of him.

There are a number of impure strains of attributionism. One could think, for instance, that while these attributions of propositional attitudes pick out the mental phenomena we are interested in, the phenomena themselves can be investigated in some other way. On another strain, one might think of our attributions as giving us only the roughest guidelines as to the real psychological phenomena they attempt to classify. Here, propositional attitude psychology would supply a somewhat idealised and only approximate guide to the nature of mental life. Lastly, and I think least plausibly, there could be a position on which one supposes that a theory of attitude ascription provides the only purchase we have on the concept of the mental, while admitting that the real nature of the mental would always outstrip our concepts of it. It is hard to make much sense of such claims for the 'noumenal mind'.

All of these positions constitute impure forms of attributionist thinking, but from here on I shall use the term 'attributionism' to denote the purest strain only.

1.3) The Three Stances

In attempting to repudiate this philosophical stance, I shall argue that facts about the cognitive psychological organisation and internal workings of an individual, brought to light by empirical research, do tell us something about the nature of his mental life and can provide more precise conditions for saying what an individual thinks and means on particular occasions. Establishing this would enable us to conclude that scientific investigations of the underlying empirical phenomena do bear on the very nature of what it is to think certain thoughts and to express

them by uttering certain sentences. Clearly, such an explanatory stance would threaten the treasured philosophical conception of these notions with which I began.

However, it might have struck some that there are other ways of denying the stance taken by the attributionist about mind and meaning without having to abandon all hope of giving a purely philosophical account of them. The thought here is that there is some fact about the matter concerning what it is for a thinker to possess a particular concept or to attach the meaning he does to his words, and that it is these facts that we are describing when we interpret someone correctly. On this conception, there is something about a subject in virtue of which an ascription to him of a given concept, or a specification of meaning for one of his sentences is true. A correct interpretation is answerable to these facts; its correctness consists in giving a faithful description of them. We could call this position the Descriptivist Stance. But now one will want to know just what facts about an individual such a priori philosophical theories are supposed to be describing? In the linguistic case, for example, what is it about a speaker that confirms that he means this or that by a given sentence? What shows that he has a structured as opposed to an unstructured understanding of a sentence? If these are facts revealed by empirical inquiry then the Descriptivist faces the same task as the Explanatory theorist of showing how such facts bear on an a priori theory about the speaker. But if they are facts to be described in a philosophical theory of meaning, or mind, then the Explanatory theorist will want to know whether they provide enough substance to answer the questions they set out to answer, while the attributionist will want to know how they can be independent of claims made from the Interpreter's stance and yet still be relevant to the facts about minds and meanings.

Although we seem to have three stances, that of Interpreter, Descriptivist and Explanatory theorist, I shall argue that the Descriptivist occupies an unstable middle position oscillating between the other two. To make out the suggestion that correct philosophical theories of mind and language should describe facts about actual language users, one would need to substantiate the claims for description as more than just the

constitutive a priori claims of the interpreter, and as less than the claims of explanation urged by the empirical scientist. I see no non-trivial way to make out this suggestion without lapsing back into either of the other two positions. In particular, where one's aim is to describe the significance of someone's linguistic behaviour, I shall argue that there is no means to correctly describe the significance it has for a speaker or hearer without offering an explanation of how they come to have an understanding of it. My brief will be to persuade the Descriptivist to chose the more explanatory option.

1.4) The Involvement of the Mind in Speech: a Prior Assumption

I shall take it as a presupposition of all parties to this dispute that there is an intimate connection between meaning and mind. To do this I shall take a little space to set aside approaches to meaning which do not accept this premiss, together with one approach which does but still proves unsatisfactory for the project I have in mind. But first of all, let us consider what it means to accept the presupposition.

The intimate connection between mind and meaning can be expressed in a number of different ways. For instance, Michael Dummett expresses it by claiming that meaning must be the object of a speaker's knowledge, and that a theory of meaning for his language is, ipso facto, a theory of his understanding. Davidson expresses the connection by stressing the inextricability of belief and meaning, according to which we cannot fix what someone means on any given occasion of utterance without knowing a good deal about what he believes, and we cannot know very much about someone's beliefs without being able to interpret much of what he says. Yet another possibility is suggested by the Gricean programme that seeks to reduce semantical notions to psychological ones. Here it is not so much a matter of an intimate connection between mind and meaning but a subsumption of the notion of linguistic meaning under mental notions. Paul Grice seeks to analyse meaning in terms of the beliefs and intentions of the speaker and hearer on a given occasion of use,

and in terms of these occasion-meanings he tries to re-construct a notion of the re-usable meanings which attach to well-worn sentences and expressions over time. In what follows I shall not have the Gricean project in mind, principally because it is an approach which neglects the structural features of a language. A word about this is in order.

1.5) Setting Aside Grice

Grice wanted to analyse a speaker's meaning something by uttering a particular sentence on a particular occasion in the presence of a hearer in terms of the speaker's intention to get the hearer to recognise that the speaker has a particular belief, and to get him to recognise this by means of the hearer recognising that this is the speaker's intention in using that sentence. Without going into detail about the further complexities of this analysis, I think we can see reasons why such an analysis leaves out an important aspect of linguistic understanding. The following example will help. We could imagine a parrot who has learned to say "Polly's hungry" whenever it wants some food and wants to draw to its owner's attention to the fact that it is in need of some food. Here, the conditions for the Gricean analysis are met. The parrot has the intention (we can say) to inform its owner of its state of hunger, and to inform its owner by means of uttering that sentence. Without stopping to argue about the plausibility of otherwise of the parrot having complex, nested intentions, we can already see that issues about how much intelligence the parrot has are beside the point. The parrot might well intend its owner to recognise its intention to inform her of its state of hunger by making that particular sound, but what is entirely lacking is any good reason to think that the parrot *understands* the sentence as a sentence of English, a sentence with a subject-predicate structure. It is this lack of knowledge about the sentence itself that should make us balk at any account of speaker's intentions as an account of the meaning of items in a public language.

It is a recognisable fact about language, seldom discussed by Grice, that different sentences uttered by speakers can sometimes exhibit common structures and parts. These facts are usually recorded in systematic theories

by clauses that specify semantic rules for the use of those recurrent parts, and formation rules for constructing complex sentences and expressions from semantically simpler ones. These rules will describe a huge variety of permissible and interpretable structures in the language. Now, without begging any questions at this stage, it is worth noting that any descriptively adequate theory of language addressing itself to the range and forms of a natural language spoken by a community of users must have recourse to something like the range of structures described by such rules. To come remotely close to describing the systematic complexity in a natural language, at the level of syntax alone, a Gricean account must make appeal to a formal apparatus that describes in detail the legitimate choices of grammatical structures which can be called upon in making oneself clear to another member of one's linguistic community. Moreover, our ability to understand novel sentences suggests that we project meanings from familiar combinations of words and structures to new combinations of those words and structures. It is hard to see how the psychologically reductionist programme can explain this. In short, the Gricean needs some account of the compositional workings of the language.

Furthermore, even if a psychological account of content takes explanatory priority over theories of linguistic content, as the Griceans tell us it does, the available choices of some linguistically complex means of communicating one's thoughts to another will, most typically, have to be explained by adverting to those structural principles. Whether or not these semantic and structural features are enough to determine fully what a speaker means on a given occasion of use, they will play an important role in identifying the sentence-type used to convey his meaning. It is more than likely that we are looking for an individuation of a sentence-type in other than purely syntactic terms. And so long as form is systematically related to content, the principles of individuation, whatever they are, will contribute to a precise articulation of the concepts which participate in determining the content of the sentence he uttered. If we can argue for such a systematic relation, the structured meaning of an utterance will be a crucial factor on which the identity of what is communicated depends.

So without yet pronouncing on whether or not a speaker's knowledge or his linguistic abilities are informed by these structural features of the language, a theory which marks the connection between meaning and the mind still has to account for the range and complexity of the language being used.

1.6) Setting Aside Model-theoretic Semantics

However we decide to construe the connection between meanings and the mind, theorists who respect the connection differ markedly from those who do not. Those who do not attempt to provide semantic interpretations of natural language by means of a logical analysis of language-world relations. Theorists of this sort view the semantic properties of a language as wholly independent of speakers of the language and their knowledge of it. I have in mind here formal semanticists like Richard Montague, Saul Kripke, and David Lewis who attempt to analyse natural language by means of model-theoretic or possible world semantics.

These formal semanticists begin by offering semantic theories for well-specified formal languages. These languages are described by their investigators as "abstract systems whereby symbols are associated with aspects of the world" (Lewis 1975 p170); although they are often related, not to the world, but to the abstract structures of mathematics. The formal semanticist constructs mappings from expressions in a language to elements in a domain. These domains are abstract model-structures. The semantic project is the construction of functions for assigning expressions in a language to elements, or the properties of them in those domains.

But we should want to ask anyone engaged in this exercise how the languages of mathematics are supposed to illuminate natural language meaning? The formal semanticist's answer is usually an argument for one of the following claims:

- (i) Natural languages (or some fragment of them) are formal languages
- (ii) Formal languages can provide a model of natural languages

(iii) Natural languages can be translated into formal languages.

Taking any one of these conclusions for granted for the moment, let us consider how the analysis works. Consider model-theoretic semantics.

Principally, semantics of this sort have to do with the *internal* logical properties of a formal language. The extensions of expressions in that language are extensions in a model, and by making assignments of the right kind to each category of basic expression, we can define the semantic values of each well-formed composite expression recursively on the complexity of that expression. However, more than one model will be available for such languages since most semantic mappings will determine the class of models up to isomorphism. In the class of isomorphic models, the extensions of non-logical constants can differ. So their extensions are not uniquely determined by the model-theory. Only relations of consequence, compositionality, and relations of exclusion and inclusion of extensions among the semantic values of expressions are fully determined. These properties provide much of what is meant on this approach by constructing a semantics for a language. The value of such an approach is usually defended as follows. The class of models for a language will place important constraints on the patterns of semantic assignments we can make to linguistic items belonging to it. This is a logico-semantic constraint on how we fix the extensions of linguistic expressions, not an account of how we do that. The only way to fix their extensions is to find some extra-theoretical means of deciding which model provides the intended interpretation of the language. This depends on what speakers really mean by those expressions. The task of finding this out is nothing short of providing an analysis of meaning for natural language sentences, and this is a project that stands in marked contrast to model-theoretic semantics. So why not embark on that project from the beginning?

I suggest, then, that these formal approaches to mathematical languages merely postpone the question of how the mind of the speaker embraces a language with those properties. At very worst, they ignore it altogether, at the cost of by-passing the project we are all supposed to be engaged upon: viz. providing a semantics for natural language. At best all

Chapter One

the tradition of formal semantics can do is to suggest that among all these (potentially infinite) possible abstract mathematical systems there are languages actually spoken by human beings. Perhaps facts about the psychologies of individual speakers would then be put to use somehow in deciding which of this huge range of potential systems were the ones that speaker's usage actually realised, or conformed to. This would be one way of redeeming the postponed promise to tell us something about actual human languages. James Higginbotham puts such a case like this:

.....linguistic theory as I have characterised it is concerned to answer the question which abstract languages are the languages of human beings, and to explain how persons are able rapidly to select a system from among those which are logically conceivable on the basis of their experience. (Higginbotham, 1983 p179)

Less psychologically relevant semantical projects would suggest that the meaning-properties of languages are independent of speakers of those languages. This project has the platonistic overtone that meanings are independent of speakers but that somehow or other, speakers have strained to become apprised of some of these standing facts. Higginbotham does not tell us anything about the actual speakers' epistemology of understanding, but he does aim to tell us how we the theorists can decide which properties characterise a speaker's actual language, assuming that we have already grasped some of the alternatives.

The less psychologically relevant project can begin with the seemingly innocent point that not everyone can know in full the meanings present in the language he uses. Consider this from Dowty et al. :

.....it is perfectly reasonable to say that no person really completely "knows" the meanings (intentions) of all of his language, despite the fact that he uses the language in a normal way. (Dowty, et al. 1981 p172)

But it often leads to the less innocent possibility that *every* speaker of a given language could be in ignorance of the meaning of some term

occurring in it. This is to suppose, as Dummett puts it, that we might "systematically misunderstand our own language" (1976, p101).

We might call this position metaphysically realist. This is the view that the properties of our languages are fixed by facts quite independent of our cognitive states. I call it "metaphysical" realism because it entails the thesis that essential properties of a language are non-conceptual, non-natural necessities. At least on Higginbotham's programme, the cognitive properties of speakers explain why we possess the language that we do, even though he believes that the facts about that language are platonic. I find both positions deeply unsatisfactory. In the Higginbotham case, the search space of possible linguistic systems would be cut down considerably, if special interest could be taken at the outset in the natural languages human beings actually use and understand. If we started with them first, using whatever philosophical and empirical means there are to fix upon and describe the actual language of a speaker, we could avoid an unguided and potentially inexhaustible search through the logical space of possibilities. Platonism is an unnecessary detour (Although it could be a consequence of one's theorising!).

In the other case the dubiously coherent possibility is envisaged that for some sentences or expressions of a community's language no member of that community understands them. This divorce between meaning and understanding should surely give us pause to wonder whether the formal semanticist was really talking about the language actually used by those speakers. If someone persisted in thinking it was, even when an account of the meanings of its expressions diverged from every speaker's use and understanding of them, one could reserve the right to by-pass such a theory and start constructing an account of what speakers actually did succeed in understanding.

I shall not consider possible world or model-theoretic approaches any further, but from now on I shall mean by the term 'semantics' the reference-specifying part of a theory of meaning for a language that also serves as a theory of what speakers understand.

1.7) Interpretation as a Philosophy of Mind and Language?

To return to matters of mind, the attributionist view I am opposing is attempting to answer a hugely important philosophical question: How can the facts about our mental lives and the meanings of our words fit into the world of facts as described by the natural sciences? While I believe this is an important project of inquiry, I do not believe that the attributionist succeeds in showing us how the material world can accommodate minds and meanings. To complete this project satisfactorily we need more substantive notions of mental states and linguistic meanings. But before I offer my alternative we have to look more closely at what the attributionist can and cannot achieve.

The best way to gain an understanding of attributionism is through the study of its methodology, and as I have indicated the guiding methodology of attributionism is **Interpretation**.

The project of interpretation sets about introducing the mind of a subject into our picture of the physical world on the basis of his observable behaviour. It aims to fix the content of person's thought and action from the beliefs and desires that give him a reason for acting. To do this, the interpreter must first describe a person's behaviour in intentional terms, explaining each episode in that behaviour in terms of a specific set of reasons.

Most commonly, interpretation works simply by citing a belief, a goal, or a value of an agent, which lies behind, in the sense of giving point to, what he does. By citing a person's state of mind we both illuminate what he or she does and explain why he or she does it. This sort of explanation lies very close to the ordinary idiom of common-sense psychology. That may be seen as being in its favour. But common-sense is not yet philosophy and in transforming it into philosophy the metaphysical picture that emerges to underpin it leaves one with a conception of the mind in the eye of the interpreter, indeterminate of content and causally inert. This metaphysical framework is far removed from common-sense, not to mention philosophically dubious. We must look at it more closely to see how these conclusions are reached, and we can do this only by examining the fundamental tenets of interpretation.

1.8) Advocates of Interpretation

As I said at the beginning, I shall be considering a broadly Davidsonian view of the mental. I say "Davidsonian " and not Davidson's view (although I won't always be so careful) because I want to render a certain conception of these matters consistent and plausible, and I am far from sure that one can render Davidson's own position everywhere consistent and plausible.

It will be quite natural, in what follows, to have in mind also the names of John McDowell and David Wiggins as exponents of this broadly Davidsonian approach. At times I shall appeal to their detailed workings out of some of the principles of this strategy, and other times to their adjustments to it.

I should also mention at this stage that by my classification, Daniel Dennett, the advocate of the Intentional Stance, also belongs to this list of attributionists. However, he can be separated off from the others according to an important distinction *within* attributionism between realists and instrumentalists about beliefs and desires. I shall say more on this later, once we have examined the method of interpretation.

Chapter Two

Interpretation and the Mental

2.1) Introduction

I shall begin by giving a sketch of the Davidsonian's impressionistic thesis about minds. This is a subtle and complex position which I cannot hope to do full justice to here; but a sketch will be enough to establish how interpretation is supposed to serve as an account of the ordinary intentional idiom. In defending this idiom from science, Davidson's position seems to owe something to a style of philosophical thinking found in the later Wittgenstein, but I shall do no more than hint at these connections. Finally, Davidson attempts to depart from Wittgenstein in thinking that the reasons for our actions are also their causes. But I shall suggest that Davidson's explanations of behaviour can provide solely rational explanations of behaviour, and this way of characterising the facts of mental life fails to fulfil a legitimate desire for a fuller explanation of those facts.

Central to Davidson's conception is the methodological strategy of interpretation. Interpretation is a form of psychological explanation of behaviour, but one which constitutes its explanandum. For interpretation does not explain behaviour construed as physical movement, but behaviour construed as action. To interpret an individual is to re-describe some of his verbal and non-verbal behaviour as linguistic and non-linguistic intentional actions, and to provide rational explanations of these in terms of a background of beliefs and desires that make sense of some of his utterances and give him reason to act. I shall now consider this strategy.

2.2) The Strategy of Interpretation

When we interpret people we see them as acting intentionally and we hear their utterances as questions, assertions, and commands. All of this reveals their purposes and to that extent their minds. But how do we do it? How do we discover the intentions with which they acted and the sense with which they uttered? This is vitally important, for however it is

Chapter Two

done, to succeed is to see other people's behaviour as the embodiment of rational minds.

Davidson will tell us that as interpreters all we have to go on is the observable behaviour of what others say and do. But, of course, behaviour as physically described is too slender a basis on which to draw conclusions about (or even recognise) minds. What we need is some way of transforming this into evidence of mental life, and we can only do this by re-describing the observable data in intentional terms, linking these descriptions in such a way as to present a rationally coherent and plausible story about the course of experience of one who behaves that way. Now as Davidson says: "we know how to do it without necessarily knowing how we do it" (1987, p441). But where we seek philosophical clarification, we need an account of what entitles us to claims to knowledge of other people's minds. And for this we need a theory of interpretation that deals explicitly with the third-person interpreter's methods for determining the contents of an agent's mind from observing his behaviour. Such a theory should reveal the nature of interpretation in general terms prescinding from the limitations of any particular interpreter. Neither Davidson nor anyone else has ever given an instance of a theory of interpretation so, just as in a style of inquiry familiar from discussions in the theory of meaning, questions are really being addressed to the *form* a theory of interpretation should take. Our reflections are philosophical ones about the value of such a conception in understanding the mental, and the chief merit claimed on behalf of this view is that it helps us to understand how mental notions can be introduced into our *descriptions* of the physical world. Put in a less bland way, the claim is that on the interpreter's view, there is no bar to our getting to know the facts about another's mind: to specify the constraints on interpretation properly and to make the right assumptions will be to make the facts about the mind public and hence accessible to others.

Chapter Two

Let me just sum up this section by reminding you of the three things there are in play:

- (i) The thinker who acts and judges.
- (ii) The interpreter who makes sense of the thinker.
- (iii) The theorist of interpretation who elaborates the principles which guide interpretation.

Davidson argues that there is no asymmetry between (i) and (ii): a thinker must also be an interpreter of others, and, since the interpreter of another must express in his own words what that other person thinks, or does or says, he must also be a language user. The theory of interpretation in (iii) is a philosophical account of what every interpreter knows. So the pronouncements of that theory must not come apart from the judgements of a well informed interpreter.

There is an important point here about the nature of mental explanation on offer in Davidson. Although a theory of interpretation prescind from the limitations of any particular interpreter's practice and addresses itself to the possibilities of interpretation in general, we must think of the possibilities of interpretation as given in the conditions of a possible human judgement (even if no-one ever fully occupies those conditions). The mental states someone has will be the ones he could be typically ascribed in interpretation that makes sense of his behaviour and conduct; and, even if they are never realised by another interpreter, the full extent of the possibilities for interpreting him will exhaust the content of his mental life.

The relation between the theory of interpretation and the ordinary interpreter depends how much of the theoretical apparatus needed to implement the former is available to the latter. The way to decide this is to produce some ground for a distinction between the *theoretical* concepts that are related to the ordinary interpreter's practices, in some way or other, from those which are *purely theoretical*, being merely artefacts of the theory that explains the methodology of interpretation. This distinction is central to the question of what is made available to an

Chapter Two

interpreter in interpretation; for an interpreter can know only what his methods and materials for ascribing permit him to know. And to the attributionist about minds, the question of what is made available in interpretation will be at one with the question of what we can know of others' minds and hence, what can be known of one's own mind:

What a fully informed interpreter could learn about what a speaker means is all there is to learn; the same goes for what the speaker believes. (Davidson 1986, p315)

So my mental states will be the states of mind a possible interpreter could interpret me as having. Indeed, as mere matters of description of otherwise physical things, they are conditional upon the possibility of giving such an interpreting description. The mind must be however it would be seen in the eyes of a well informed interpreter. So it becomes vital to know what interpretation can provide of the character of minds.

Before we can even begin to answer the questions just raised we have to look more closely at the constraints and assumptions that are at work in interpretation theory. This is what we must now do.

2.3) The Principles of Interpretation

Rationality plays the key role in Davidson's thinking about the mental. It guarantees holism and the irreducibility of mental notions. We shall come to these consequences in a moment but let us focus on the operative role of rationality itself.

When we re-describe others' activities in intentional terms, rationality provides an a priori link between beliefs, desires and behaviour, described as intentional action; while other a priori links trace out the rational ties between beliefs and further beliefs, desires and further desires. All of these connections are drawn upon when we begin to interpret others. At the outset, we have to find reasons for what they say and do given in terms of what they think and what they want; and in doing so, we must also supply reasons for their having certain beliefs and desiring certain outcomes. For if we treat these beliefs and desires as the springs of their actions but find

Chapter Two

no reason to think they have such thoughts and urges, or worse feel that there is reason to think they lack them, then we have failed to explain their acting intentionally in this instance. So there is a constraint to be given here on interpretations: the intentional re-descriptions the interpreter gives to behavioural events must satisfy the a priori principles of rationality.

These rational links are operative within the mental scheme, they relate complexes of intentional notions so as to meet:

....conditions of consistency and rational coherence [which] may be viewed as constitutive of the range of application of such concepts as those of belief, desire, intention and action" (Davidson 1980, p237)

This does not tell us how to apply those concepts in actual cases, nor in general terms does it relate those concepts to physical conditions, but it does place a constraint on any putative re-description of a behavioural event. In re-describing such an event, it does require us to relate the mental concepts we apply to someone's behaviour to a range of other mental concepts to which it is rationally tied. A word about this is in order.

Attitudes like belief and desire are identified in part by their propositional contents. These are the propositions towards which the attitude is directed. The logical relations among these propositions account for many of the links in the network of attitudes. One cannot have (be ascribed) the belief that Mary lied in court without having (being ascribed) the belief that Mary spoke in court, the belief that there are such things as courts, and so on. Conversely, one cannot have (be ascribed) the belief about Mary lying in court if one also has (can be ascribed) the belief that Mary did not speak in court, or that courts are places where the use of language is forbidden. The grounds for attributing one belief include the grounds for attributing many others, to which it is logically related through its object, and which help to determine its content. The same is true for desires and other propositional attitudes. Meanwhile beliefs and desires are connected to one another and to actions by the principles of

Chapter Two

practical reason. To act intentionally is to do what one believes will achieve what one most wants at that time. Correlatively, if one believes things are not as one wants and one has no conflicting beliefs or countervailing desires one has a reason to act.

A theory of interpretation that provides redescription of people's behaviour has, for any particular case, to massage these two requirements of reason into a huge interlocking system of beliefs, desires, intentions and actions that explain the reasonableness of the agent's behaviour given the rational spread of his mind. He must be shown as acting in accordance with his thinking on the basis of the things he wants. So we understand people's minds on the basis of their attitudes and actions; attitudes and actions we attribute to them as interpreters. And:

...if we are intelligibly to attribute attitudes and belief, or usefully to describe motions as behaviour, then we are committed to finding, in the pattern of behaviour, belief and desire, a large degree of rationality and consistency" (Ibid)

Thus when I interpret someone's behaviour there is always a presumption of rationality in his or her favour. This means that if we find someone wanting in reason under one interpretation, we should try to replace it in favour of a more penetrating description that preserves that person's rationality. For it is only by finding a satisfactory set of reasons that I can find the behaviour and speech of another so much as intelligible, and only in so far as someone is rationally intelligible that they have a mind at all. This is because Davidsonians believe that the giving of such rational explanations is inseparable from the claim that the person explained has beliefs, desires, hopes, fears, and a set of values he can act on. It is only when certain behavioural events stand in need of rational explanations that we detect the presence of minds. In the physical world, rationality makes room for minds: it allows us to gain sufficient distance from the physical events by allowing us to describe the rational structures of embodied minds:

In order for the intentional notions to make sense we must require enough rationality to let our pattern of explanation be reason

Chapter Two

explanation rather than merely causal explanation. (Follesdal, 1982, p312)

Causal explanations will still be operative but they will subserve rational explanation. (I shall have more to say about this in a moment.)

Notice that action must be included among the intentional notions defined a priori within the structure of the mental scheme. Therefore the mind stretches, on Davidson's picture, to include actions. Actions are observable moments in people's mental lives that can also be described as events in their physical histories. This provides Davidson with an observational foundation to mental notions and hence facilitates his claims to knowledge of other minds. But notice too, that because we must work our way back from these observable points of entry on mental life into the interior of the rational mind to cite the reasons for action we are also giving an interpretation of someone's thinking. We use rationality to trace the links between someone's actions and his thoughts, his wants and wishes, his hopes and fears.

In discussing rationality we should also note the force of its constitutive claim on us, as thinking beings. We have acknowledged that when and only when we are entitled to give explanations that treat a creature as embodying a certain rational structure, can it be said to have a mind. So when all attempts at providing reason-giving explanations of a person fail, we must conclude that there are no grounds for ascribing our subject any intelligible thoughts at all. The loss, or absence, of rationality in you would lead me to conclude that you had literally lost your mind. Only to the extent that I can treat you as rational, by sustaining a rational explanation of you, can I find any grounds for your having a mind at all. We now have an explanation for the presumption of rationality. We expect to find people rational because those are the norms that interpretations of their behaviour must conform to; and to have a mind, or be an agent, is to be a proper subject for interpretation. So minds will conform to rational standards just in case interpretation is governed by rational norms. The interpreter will also try to see his subject as guided by the principles of rationality, but that is because those are the principles that guide the interpreter in giving interpretation to his subject's behaviour.

Chapter Two

There can be no gap between our knowledge of other minds and our way of understanding them. The interpreter will view the subject as a rational agent and will even try to capture that agent's rational outlook on the world, but all of this will be constituted by an act of interpretation. This is the short step the attributionist needs to the claim that rationality is constitutive of mental life. It tells us what minds are. As Michael Root puts it, in paper that gives a very sympathetic and favourable reading of the line taken here:

...we owe our idea of the mental to our interest in explaining the behaviour of others, and...our idea of the mental is constituted by the way we pursue that interest: we offer a rational explanation of the behaviour. (Root 1986, p294)

The accommodation of the mental in the physical world is then aptly expressed thus:

Other minds, on Davidson's view, are what we get when we interpret the behaviour of others. Bodies are what we have before we interpret their behaviour. (loc. cit)

The constitutive principles of mind are the a priori norms of interpretation. They are normative because they say what a person ascribed certain beliefs *should* think, and what someone with those beliefs and accompanying desires *should* do; if they do not, the grounds are inadequate grounds for their attributions. The principles are a priori because the interpreter does not contemplate people's thoughts and actions and measure his principles against these, instead he measures the claims that people have thoughts and actions against the standards of interpretation. Again, this adds up to the fact that the interpreter is telling us what minds are.

Chapter Two

2.4) The Holism of the Mental

Another feature of the Davidsonian view is the holism of the mental. We have seen how the attribution of one belief requires many others, we make sense of beliefs only in so far as they cohere with others, and rule out incompatible beliefs; the same for desires, and other attitudes. (Interestingly, hopes and fears are often said to be less rational.) Moreover, the beliefs and desires we cite in explanation of a person's action must belong to this interlocking system. No simple belief-desire attribution will serve the needs of interpretation on its own: it must conform to other interpretive acts and it will impose requirements on them. These interconnections are assumed as we begin to interpret others:

Beliefs and desires issue in behaviour only as modified and mediated by further beliefs, desires, attitudes and attendings, without limit. (Davidson 1980, p217)

Particular interpretations of events only serve in the light of all others because their grounds depend on much else that goes on in the subject's mental life, elsewhere and at other times. So:

..we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions and decisions. (Ibid., p221)

Even if the details of such a framework are never fleshed out by an actual interpreter, they will always be assumed. The assumptions will not consist in implicit assumptions about the background theory but in an explicit assumption of the interpreter that the elaborations of the background thinking of him and his subject are largely the same. Davidson calls this the Principle of Charity:

No simple theory can put a speaker [agent] and interpreter in perfect agreement, and so a workable theory must from time to time assume error on the part of one or the other. The basic

Chapter Two

methodological precept is, therefore, that a good theory of interpretation maximises agreement. (Davidson 1984, p169)

But even where error is found, in the guise of false belief, the well informed interpreter will still find it explicable error. It is in the same way that a creature does not have to be perfectly rational either. This is important since otherwise it could be assumed that it counts as an objection to the Davidsonian conception of mind to point out that people are not always rational. However, the interpretative approach does not require people to be perfectly rational; it will simply insist that *irrationality* only makes sense in terms for someone's going against a norm or standard already operative in that thinker's mental life. In this way we can contrast these failings of rationality with the cases where the standard of rationality has no application at all (infants, animals, the insane). These are non-rational animals whose behaviour is to given a causal explanation, presumably.

But how much rationality is needed to insist on more than a causal explanation of a creature's behaviour? This is hard to say. But Davidson's use of the conditions of rationality looks rather too strong, and it seems reasonable to think there might be room here for more flexibility. Davidson thinks that in ascribing beliefs and desires, and other propositional attitudes, to people we must respect the logical and semantic properties of their propositional objects. Thus we should expect a good deal of consistency (logic) and widespread agreement with the environment (semantic) in the set of attitudes we attribute to a person. So in interpreting an agent in conformity with Davidson's Principle of Charity, we must maximise the number of truths the agent believes and expresses, and make him come out as rational as possible. This seems too strong, and others (Grandy, Wiggins, Follesdal) have suggested a less strong but more discriminating Principle of Humanity. There are many versions, but a good statement of what is required is given, here, by Follesdal:

When ascribing beliefs, desires and other propositional attitudes to a person on the basis of observation of what he does

Chapter Two

and says, do not try to maximize his rationality or his agreement with yourself, but use all your knowledge about how beliefs and attitudes are formed under the influence of causal factors, reflection, and so forth, and in particular your knowledge about his past experience and his various personality traits, such as credulity, alertness, reflexiveness etc. Ascribe to him the beliefs and attitudes you should expect him to have on the basis of this whole theory of man in general and of him in particular. (Follesdal 1982, p315-6)

Rationality, amongst other things, tells us that when we attribute beliefs, desires and the other propositional attitudes to people we must be careful to observe agreements and relations between them, to ensure that the set of attitudes we are ascribing is consistent and sufficiently inclusive to make sense of the attribution of any single attitude. This is what led us to recognise the holistic character of interpretation. And as we now know, it is a short step from here to the holism of the mental: an event is intentional only as described, intentionality is the mark of the mental, interpretation provides intentional descriptions of events holistically, and so the mental is holistic.

It is this holism and the norms of rationality that give rise to it that together frustrate any reduction on the intentional level to the physical level, and dictates a holistic fit with the behavioural evidence. Fitting an intentional description to a behavioural event requires us to locate that description in a swirl of others in the mental scheme first, before determining whether the total interpretation can be fitted to the behavioural facts. The fit is holistic, but more importantly, irreducibly holistic.

Now having said that, there must be some principles linking interpretations to behaviour because otherwise the physical circumstances would provide no constraint and all interpretations of people could be the same rationally coherent description. Furthermore, we could not check the fit, even holistically, if we could apply re-descriptions to just any piece of behaviour. There are certain minimal conditions to be met before a piece of behaviour undertaken deliberately by a person could count as her deliberately slapping his face. The needed links will be a second set of a

priori principles that connect the intentional level with the physical facts, in the sense of constraining what can count as a behavioural implementation of intentional action.

Together these two sets of a priori principles must pronounce what it is for an interpretation to be correct. This is to be distinguished from pronouncing that an interpretation is the right one. There may be more than one correct interpretation of an individual compatible with all the behavioural evidence there can be. In these cases, neither the principles nor the behavioural evidence can adjudicate, and these are all we have. So in these cases, there is nothing to choose between them.

Finally, notice that Davidson's conception of mind provides further illumination of the ultimate a priori status of the project of understanding the mental mentioned at the very start of this thesis. There I said that the a priori intuition was that because we know other minds, we understand what minds are. This follows from the present conception because minds are what we understand when we interpret people's behaviour. Philosophical investigation can then illuminate theories of interpretation that tell us what we understand when we interpret others and how we come to do so.

Now that we have enough of an overview of this position, we must turn to some of the difficulties with it and the necessary refinements that must be made to it.

2.5) A Third-Person Perspective

At first, what is so surprising about Davidson's position is that it assigns privileged access to mental life to the third-person, unlike the more familiar Cartesian pictures of the mind on which first-person experience secures certain knowledge of our own thinking. Davidson believes that the mind is best known from the third-person standpoint. In fact, for Davidson, this is the best way to secure knowledge of the mind. But this interpreter-relative conception of mind might seem to miss some essential features of a more agent-centred view. By making minds depend on facts about the materials and methods of the interpreter, it is in danger

Chapter Two

of squeezing out the access we have to our own minds. Davidson is not unaware of having made knowledge of one's own mind more problematical than knowledge of other minds, but he is given to say:

sometimes I learn what I believe in much the same way someone else does, by noticing what I say and do. (Davidson 1987, p441)

What are we to make of this? Surely, this perverts our conviction that we have immediate awareness of ourselves, and distorts the facts about self-knowledge? The reply would be, not necessarily. When I suppose myself to hold a particular opinion, or to have a particular belief, I am subject to the same standards of interpretation I would apply if I ascribed this belief or opinion to others. I cannot attribute one belief to myself without supposing there are a great many other things I am thinking besides: individual beliefs depend for their content on many others. So I cannot believe that that Mary lied in court without knowing something about the practice of courts and truth-telling. If a body of beliefs I hold is recognisably incoherent, I must give something up or abandon my reason. When I change my beliefs I should be able to offer a reason, when I act intentionally, I must have some view of my purposes. And over time, I must find enough continuity in my mental life to be able to understand myself fully at all. There are limits on how adjustments of my view of myself are to be made, and limits too on how accurate I am. So when I attribute belief to myself, the aptness of my attribution is judged by the same standards I use to judge the aptness of making belief attributions to others. For the Davidsonian, the difficulties to be overcome on the first-person case are not as daunting as they first appeared.

2.6) The Narrative Character of Interpretation

Having a mind is a matter of making sense, but making sense to whom; to oneself, to others, to all others? Davidson would try to ease the idea of any discrepancy between these answers. One can only be made sense of by anyone because one makes sense in general. But we have to take the locution of "making sense of someone" very seriously on the

attributionist's picture. We do not discover sense in people, we *make sense* of them; or if one prefers, we discover the sense that can be made of them. (Consider the Glaswegian reply to the accusation, "Are you trying to make a fool of me?" which runs, "You don't make fools you discover them." Something of the force of this reply is lost on the attributionist.) Whatever the talk of discovery comes to, the attributionist will insist that it is only in so far as there is a way of rendering someone's thoughts and activities intelligible that they can be taken to exhibit mental life, or fulfil the conditions for thought. Having a mind consists in one's behaviour being intelligible from the standpoint of a judging subject who can provide rational explanations of it. So having a mind is a matter of what is available in interpretation. But now we need to know by which interpretation(s) one makes sense. By what enlarged, or restricted standard is one rational? Is there more than one way to construe the nature of interpretation. The answer to this will be yes if we compare the ideas of Freud and Davidson. What is illuminating here is not the differences between these accounts of mind but their similarities in appealing to the third-person methods of interpretation.

To interpret a person's actions is to provide reasons for her doing what she does; it is to ascribe beliefs, desires, wants and wishes that make it reasonable, by her lights, to do and say what she does and says. As an interpreter, I express in my own words what she is thinking, and I cite particular beliefs or values as an insight into the motives for her behaviour. I report indirectly what she thought and intended, and what she meant by what she said. And by giving the background of thoughts and feelings which impelled her at any given moment I am describing the reasons for what she does. But as an interpreter, I will always try to make sense of someone in my own terms, which means trying to find them rationally explicable by my standards. But I may not have the terms to capture subtleties in another person's thinking, and I may find her irrational by my standards. Have we the right to conclude that there is no sense to be made of her, and that in failing, by my lights, to be in any way rational, she lacks a rational mind and the capacity for thought altogether? Surely not. We cannot draw conclusions from my limitations to the limits

Chapter Two

on her susceptibility to interpretation of any kind. It depends how far we look for an explanation. The point requires some subtle handling.

Davidson will point out that it is not just up to me to describe the contents of someone's thoughts, the meanings of her words, or the intentions behind her actions in any way I choose, and yet how far should I go, or how soon should I refuse to stop in trying to give reason-guiding explanations of her behaviour? To interpret someone correctly, I must be guided by certain standards of interpretation. As we have just seen, a presumption of rationality is one of those standards. But it was also said that I interpret someone by judging her thoughts and actions rational by *my* standards and in *my* terms. So the question remains: why should my standards and my terms be thought to be the right ones? Davidson's answer to the first of these is going to be that in so far as I am a rational being at all, my standards just are *the* standards of rationality: the standard that applies to all humanly intelligible creatures. But on the second point, Davidson will concede that he is not really concerned to examine the peculiarities and limits of what is available to any particular interpreter, but is more concerned with the conditions for interpretation in general. The possibilities of interpretation, if they can be brought to light, will provide the constitutive conditions for introducing minds into the world. The possibilities will be given in line with the general principles of interpretation. The principles are constitutive of the mental because we can only regard as mental what conforms to these standards. But have we got any very precise idea about how these standards apply, and where we are being too charitable in interpreting someone, or not insightful enough?

This is vitally important since success in interpretation makes for minds according to Davidson. Illumination of both my mind and the mind of another can be given by concentrating on the conditions for making sense in general. I can understand another only in so far as I can make sense of him, but my being able to do so supposes that I have the means for making sense. But what if I lack the means for making sense of others who can nevertheless be found intelligible in some way. Is this just a failing of moral imagination on my part? Will just any way of

attempting to make someone intelligible do? How far should we go? Interpretation theory is supposed to tell us the answers to these questions, but remember that interpretation theory must operate within the sphere of possible human judgement providing us with a way of examining our apprehension of others' minds. We need some ground rules here because it is only by examining the way in which sense can be made of a subject's experience and behaviour that the conditions for having a mind can be brought to light. If there are no interpretation-independent facts to keep faith with, do all convincing reconstructions of someone's thought have a claim to authority? We need to consider this in the light of what I want to call the narrative character of interpretation.

It should now be easy to see that on one reading of him, Freud's psychoanalysis is a form of the Davidsonian method of interpretation we have just been rehearsing. Here too, there is a desire to ascribe reasons to people to explain their behaviour, even if underlying ones. The searching out of reasons in psychoanalysis attempts to restore a kind of sense that can be made of rather aberrant behaviour. It searches for a coherent narrative that makes sense of an individual while being unique to that case. In Charles Dickens' Great Expectations, we confront the sorry sight of Miss Havisham. She has been jilted, she will never recover and cannot accept that her bridegroom has deserted her and that the marriage did not take place. But as she sits in her wedding gown, she is not waiting to be wed, she is wasting away; not accepting, because unable to, the humiliating rejection that denied her love. We tell the story about her experience, about her unique cast of mind, and we say, "Now do you understand why she sits there in her wedding weeds: life has deteriorated for her, she has resigned herself to decay".

We provide an interpretation of Miss Havisham in order to understand her strange behaviour, and by giving an interpretation we can make sense of her. I think this brings out the narrative character of the interpretive stance. But is interpretation just a narrative response to the desire to understand someone's behaviour? Might there be many such responses? In being externally imposed there is a danger of fictionalising the facts about another's mind. If Davidson is going to give the central role

Chapter Two

in mental epistemology to the interpreter, he has to give us some reasons for thinking that what an interpreter offers us in his descriptions of another's mind is in fact correct, that it captures how things are with that agent.

Is there any substance to the thought that a correct interpretation is also an explanation, and that it may or may not correctly explain the course of someone's behaviour? Well, it is of course a rational explanation: it has coherence, consistency, it is a plausible account of a mental life. But is it the correct account for this or that person? What does this mean here beyond satisfying the internal standards of the theory?

A natural thought would be that there must be something about the agent herself in virtue of which the correct explanation is true. To opt for this intuitive suggestion is to turn interpretations into something more like hypotheses, where we suppose that the theory-independent facts about minds are there to be described by the correct interpretation; they are facts against which interpretations must be checked. But this is just what the interpreter denies, and his denial is the whole thrust of his case.

Someone might object that the attributionist can accept that there are some facts which amount to the evidence for interpretive claims; although these would have to be facts about behaviour. But this is unsatisfactory: either such details have already been taken into account, in conferring a *prima facie* plausibility on a suggested interpretation that relates intentional descriptions to the prevailing behavioural conditions; or else they are no more than the bare behavioural conditions, described in such a sparse, physicalistic way that they could not support particular attributions at all. Either way, Davidson, qua attributionist, is not permitted to entertain the intuitive picture of a set of independently constituted facts to which a correct interpretation is answerable. But what other view of explanation can he give? What other way is there to confirm that there is something right or wrong about a cogent description of another person's thinking? The attributionist will see these persistent questions as misconceived: what we must grasp is that the internal standards on the rational coherence of any narrative interpretation

Chapter Two

provide the only view of what counts as a correct explanation of someone's behaviour.

Yet still there is a strong inclination to ask, what is it that makes us so successful, when we are, at giving sustained interpretations of people and the language they speak? Failure to answer this leads us back to the fertile analogy with the literary critic who goes on providing a number of penetrating interpretations which may remain illuminating though external to the mind of the author. I will return to this reading of Davidson and Freud, but it is useful before doing this to remind ourselves of the deep reasons why Davidson refuses to compare the rationally coherent narrative of an interpreter with any confirming facts outside the interpretation. It is here that it is worth drawing out some of the Wittgensteinian themes in the Davidsonian conception.

The reason for failing to look for further justification springs from the thought that there is nowhere else to look for facts about the mind than in the grounds given by the rational interpreter.

The difficult thing here is not, to dig down to the ground; no, it is to recognise the ground that lies before us as the ground.
(Wittgenstein 1956, VI-31)

and so we have to resist the temptation to dig down below that level in an attempt to find some ultimate justification for our claims. Wittgenstein calls the level at which justifications come to an end, 'bedrock'. This is the level at which people can be seen as behaving intentionally and using language correctly. It is the most basic level at which minds and meanings can be accommodated. Wittgenstein warns us not to dig below bedrock to a level at which everything normative disappears from the picture. So for Davidson, behaviour stripped of its intentional gloss consisting of bodies that issue physical movements and utterance-sounds, and perhaps have dispositions to behave, belongs below bedrock. No doubt these events are related in interesting ways to what can be seen implementing at the level of bedrock, but we can make no use of these relations nor gain any insight into how things stand at the higher-level. So it is only behaviour perceived as the intentional action and speech that constitutes the level of

evidence that supports claims for particular attributions; but now we can see that this sort of behavioural evidence belongs within interpretation, it is a constitutive part of an interpreter's grounds for particular ascriptions. It is only with reference to behaviour construed *at the level of bedrock* that we can claim that the mind of another is revealed to a fully informed interpreter on the surface of his practices. Even the Freudian could be accommodated on this picture of bedrock and the surface of a practice. Treating the Freudian picture in this way would introduce cases where the possibility is left open that the analyst who supplies an interpretation, but not the patient himself, has access to those facts. What we need here is a picture in which what is "unavailable" to the subject can be brought to light by the analyst by his finding the right means to describe the patient's behaviour. The "inaccessible" intentional notions would still belong to the surface of behaviour, though exactly how they applied to it is something not known immediately but taking a bit of working out.

To accommodate Freud and Davidson, we see that the notion of the surface of an agent's practices must be behaviour characterised in a richer vocabulary of intentional descriptions. The surface must be seen by the psychoanalytic or the rational interpreter as exhibiting the purposes of others; it is not viewed by us, except with conscious effort, or determined resistance, in the thinly described way offered above. The term "behaviour" is simply ambiguous as between mere physical bodily movement on the one hand, and intentional action on the other. Of course, the former coincides with the latter at a time and place, and the former implements the intentions of agents revealed by the latter. So bedrock for us is the level of intentional behaviour of agents whose exact motives have to be worked out in accordance with the principles of interpretation. At bedrock, we have to treat the things people do, the bits of their behaviour, as intentional, done for a reason and motivated by the things they think and the things they want. We may describe people as performing certain kinds of action, but in any particular case, our right to that claim, and to the interpreting description under which we see the agent as acting intentionally, has to be won by elaborating more of his or her mental life.

Chapter Two

What lies open to view on the surface cannot be illuminated in any way by facts below that surface: below the level of bedrock. The possibilities of giving genuine explanations are limited, but according to Wittgenstein, we are none the worse for that. In fact, he never tired of reminding himself that the need for an explanation in certain cases was misconceived: "for someone worried by love, an explanatory hypothesis will not help much".

However, it is difficult to give up the thought that the sub-bedrock facts are intimately related to what goes on at the level above. As John McDowell puts it it is hard:

to avoid acquiring a sense of what, as it were, lies down there: a web of facts about behaviour and 'inner' episodes, describable without using the notion of meaning. One is likely to be struck by the sheer contingency of the resemblances between individuals on which, in this vision, the possibility of meaning seems to depend, and hence impressed by an apparent precariousness in our making sense of one another. (McDowell 1982, p348)

While agreeing with McDowell that it is hard not to acquire a conception of what lies down there, it is hard not to acquire a better conception than this. It depends where below bedrock one is operating. From detailed empirical investigation of the cognitive organisation of language users and perceivers, it is less easy to be struck by the sheer contingencies. It is only by looking at the level of physics or neuroanatomy that this phenomenon and the sense of precariousness re-enters. In the second half of this thesis I shall suggest a few ways in which we can gain not only a conception of how things stand below bedrock, but also some knowledge of the connections that must hold between lower and higher levels to sustain complex intelligent activities like one's mastery of a natural language. But this must wait.

Returning to Davidson, we confront the following problem:

an event is an action if and only if it can be described in a way that *makes it* intentional. (Davidson 1980, p229, emphasis mine)

Chapter Two

But does he himself have anything more to say when faced with a number of alternative narratives each offering to supply the agent with different reasons for action, giving more or less penetrating descriptions of what he is up to? Turning to the agent himself, as the Freudian will know, may prove of limited value. Davidson notes that:

As observers we often describe the actions of others in ways that would not occur to them. This does not mean that the concept of intention has been left behind, however, for happenings cease to be actions or behaviour only where there is no way of describing them in terms of intention. (Ibid.)

But this does not help us with our question about which way we should describe someone in intentional terms. As it turns out Davidson does have more to say, and as it also turns out, he takes the same course in narrowing the gap between narrative interpretation and explanation of behaviour that Freud took. There are no interpretation-independent facts about the *mental*, but there are non-intentional facts about the *causal* history of someone's behaviour. This is the step that departs so radically from Wittgenstein, ignoring his advice never to confuse, or conflate, a conceptual question with a causal one. Wittgenstein's penetrating remarks about Freud might seem at first sight to carry over to Davidson. But I shall argue that Davidson does not succeed in intertwining the rational and the causal explanations to any effect and that what we are offered in the end are just narrative descriptions (rational explanations) of the mysteriously regular but detached events in the causal swim. These must take a course of their own according to Davidson, unaffected by the stories we tell ourselves afterwards.

Chapter Two

2.7) Rational and Causal Explanation

Both Davidson and Freud would accept that to make sense of someone we must find application for a range of attributions of thoughts and reasons for actions, and these interlocking beliefs, desires, intentions and actions must conform to some sort of rational pattern; for these are the patterns of the rational mind. (For Freud, the 'rationalising' mind, with an intended pejorative tone.) Freud will attempt to draw on more material in describing these patterns; he will seek underlying reasons which help us to make some sense of the seemingly aberrant behaviour of others; he will find a kind of intelligibility here, where the Davidsonian can find none. But in all cases of action, Davidsonians like Freudians will want to know the very motive on which a subject acted; in Davidson's terms, we want to know the primary reason for acting. As the Freudian points out this is not always clear. But nor is it entirely unproblematical for the Davidsonian. Davidson offers us the case of Oedipus who had a reason (a good reason) to kill his father, but who does not act on that very reason when killing Laius on the road to Thebes. One may have a good reason that *would* explain the action and yet that reason might not be the reason the subject did it. So giving plausible explanations in terms of actual reason is one thing, knowing on what basis the agent actually acted is another. But what force is there in this thought that there is some difference here? What factor makes one reason operative rather than another? The answer Davidson will give is based on causality. Reasons bring about the behaviour they explain because reasons are rational causes:

So when we offer the fact of the desire and belief in explanation, we imply not only that the agent had the desire and belief, but they were efficacious in producing the action.....the belief and desire were causal conditions of the action. (Davidson1980, p232)

Yet a desire and belief might be the reason that caused Oedipus's action without being sufficient, ie. explanatory for the correct interpretation: Oedipus might be intent on cutting down every man in his path until he

Chapter Two

gets to his father. What Davidson needs to say is that to be the actual reason that explains a particular action, it has to be the one that causes the behaviour *in the right way*. This leads to problems we must consider a little further.

The narrative character of interpretation plays a crucial role in fixing the nature of the mental. Someone has a rational mind when there is sufficient coherence in her activities to sustain a plausible description of her mental life from the interpreter's standpoint: we introduce the materials of mind when there are grounds for giving a rational explanation of the person's behaviour. But what we want to know is whether it is the right explanation of that person, and where this is a genuine choice, Davidson will say the deciding factor is causality.

But the causal factor, if it can be made out, will take us beyond the purely narrative conception of interpretation; some accounts but not others will chart the causal course of events. This should be seen as a good thing because the narrative conception threatens a deeply held intuition about ourselves, namely, that we are the originators of our own meanings and purposes. Were we not the authors of our own actions, and were the facts about our minds entirely answerable to the decisions of an interpreter, or exhausted by the possibilities of interpretation, we would be left with a picture of the mind in which, like Hamlet, one would be in a play full of purpose but with no purpose of one's own. Whereas to align interpretation to the causal facts is to return some of the motive force to the agent. This is not to deny attributionism, however. Rather, the idea is that by linking the subjective conditions for thought about the mental to the physical causes of agents' actions we give a complete account of the objective reality of minds. The role of the interpreter is ineliminable - to lose sight of this is to lose the observable foundation for knowledge of other minds - but what the interpreter picks out is a feature of the agent himself: a causal feature. We must now find a way to connect the third-personal methodology with an ontology of inner states.

The obvious place to start is with action. Actions are episodes in an agent's physical history in which he succeeds in embodying his mind in his behaviour. But as events in the physical world, they have physical

Chapter Two

causes; while as bits of intentional behaviour we think of them as caused by mental states. To avoid over-determination of these events we must find a way of tying together the terms drawn from different types of explanation: the rational and the causal. Davidson supposes we can do this on occasions by treating the intentional and the physical terms as co-extensive: reasons are causes. The physical states which cause behaviour must, at the same time, be intentional states if descriptions of them are to enter into rational explanations of behaviour. And it is only when intentional descriptions of physical states can serve in rational explanations that the behaviour they are causally responsible for can be deemed intentional action. Only then can we say that intentional states cause actions: reasons are rational causes.

Here we have a picture in which, for any behavioural event, it is describable as an intentional action if and only if there is a reason which is at the same time its cause. Reasons we may supply might seem to cast explanatory light but fail because they are not really the causes; while the causes do not explain unless they are also the reasons. Nevertheless, being the causes of behaviour does not make them the reasons they are, and being the reasons they are does not make them causes. An intentional event causes behaviour (physically described) in virtue of being a physical event token identical with it. It is as a physical event that it falls under causal laws. What we need for these causal sequences to subserve rational explanation is a legitimate way of identifying a reason with a cause. This is important to the Davidsonian because a reason is efficacious if it is the cause of action, and explanatory only if it causes in the right way. This idea marks Davidson's departure from the Wittgensteinian thread teased out above. Wittgenstein warned against conflating conceptual questions with causal ones. He would have denied that there was any way of tracing the real reason for someone's acting as they did back to the way it was caused. Now in the light of this Wittgensteinian criticism we should do well to ask whether Davidson is entitled to the idea as he expresses it.

How is Davidson to account for the efficacy of particular mental states? He tells us that "when events are related as cause and effect, they have descriptions that instantiate a law" (Davidson 1980, p215). So do

Chapter Two

psychological states ever instantiate laws? The answer is: not under their psychological descriptions. For Davidson, there are no purely psychological laws, because psychology is not a closed system of description. It relies on many physical conditions which could not themselves come under psychological description. This just leaves open the possibility of psychophysical laws, but as we shall see Davidson thinks there are no laws of this sort either because of the irreducible and normative elements in psychology. The only strict laws are physical laws, those couched in purely physical vocabulary. So it is only when a pair of events instantiate the antecedent and consequent of a physical law that they are related as cause and effect.

In the light of this, how can Davidson equate the efficacy of a reason to act with that reason's causing some intentional behaviour *in the right way*? To succeed here, he would have to select the physical underpinnings for the belief and desire states that serve as reasons and then assess them in some way. But how can this be done? How do reasons direct us towards physical causes? Surely, the only way to identify reasons with causes is ineluctably holistic: it is a matter of finding it intelligible (reasonable) to give an event an intentional description that belongs to an interpretation scheme that has application to all other events in the agent's behavioural history.

It is hard not to acquire a conception here of different types of explanation running in parallel: there will be patterns of reasons and actions rationally connected at one level, and neural and behavioural events causally connected at the other. But parallel lines do not meet and it is difficult to see how when two different reason-giving explanations can be given roughly equal weight in accounting for someone's behaviour it is supposed to be a causal matter which of them is actually responsible for the outcome. Instead it is hard to resist the idea that the causal sequences of physical events would have been just as they always were anyway, whichever reason was operative, and that it is just a matter of choice, within principled limits, of what rationally coherent narrative we provide as an intentional gloss on those events.

Chapter Two

On this conception, the intentionality of a mental state is causally inert, it merely accompanies the physical facts in an epiphenomenal way. Reasons cannot cause in virtue of being the intentional states they are, they can only be applied as descriptions to physical events that cause. What Davidson needs to say is what it is to be able to describe a cause as a reason. More precisely, by finding the conditions under which we can interpret people's behaviour as intentional action we thereby earn the right to re-describe the causally responsible inner states of them as the beliefs and desires that provide the rational explanation of the causally produced behaviour. But of course, as I have said, the conditions for interpreting an agent in these rational terms depends on the propriety of attributing to him all the surrounding propositional attitudes that it is intelligible that he should have these reasons, and this can only be determined in the light of his behaviour. Once again holism and rationality show that it is the totality of intentional descriptions of someone that license the attribution of a single reasonable action. So the causal feature of the event will not fit with the relevant portion of mental life in isolation from all others. Any description of causes as reasons demands many other interlocking reasons subserved by other causes. The fit is always holistic. What we need to know is when are we entitled to apply a total scheme of intentional re-description to a physical creature? The inner physical states of agents causally explain their behaviour described in physical terms, and beliefs and desires rationally explain their actions described in intentional terms. But when can we apply one scheme to another? Not all human beings will turn out to be minded, but where they do, we want to know why, and this question goes beyond the question of what is it to have a mind. In effect it is to ask, why are we successful when we are at describing people's behaviour in rational terms? We have failed to discover any more tangible way of intertwining the rational and the causal schemes of explanation, operating at the intentional and physical levels of description, respectively, than by applying a total interpretation scheme to a physical creature. So decisions about a person's reasons for acting will be referred to the total schemes of interpretation. Within any one scheme there should be a way of resolving competing

Chapter Two

explanations that rationalise the person's behaviour, and where alternatives persist, the difference marks a boundary and a choice between interpretation schemes. There can be no overdetermined outcomes in intentional behaviour as far as the interpreter is concerned. Our knowledge of people's reasons for action and hence our knowledge of their minds is a matter of the choices between interpretational schemes. The extent to which there is a genuine choice between them and we can choose between them, is the extent to which we can know others' minds. There is no more to the mental than the fully informed interpreter can know; so psychological states are constituted in accordance with the *a priori* assumptions that guide us in describing them and the conditions that entitle us to apply our descriptions in actual cases.

For the Davidsonian, the intentional level is somewhat aloof, and removed from the causal facts. That being so, he is required to regard psychological explanation as a form of narration. This is what interpretation of someone's mind, speech and action amounts to: a superimposition of an intentional network of propositional attitudes, speech acts and actions on the vast repertoire of a person's behavioural output. Of course such an intentional re-description must be constructed within principled limits and be subject to constraints of application. But if we cannot find grounds in the internal states of the agent for his being a believer or desirer of this or that, if there are no physical, neurophysiological or purely behavioural grounds for applying those terms to him, then we must treat it as only matter of our being able to describe him intelligibly as a believer and desirer that he has mental states with propositional contents. The form of our propositional attitude reporting sentences will be the form of his thought. But the causes of his behaviour will give no indication of this in themselves.

Is this the best we can hope for in terms of combining a causal thesis with the psychological explanation of action? Davidson would say 'yes', and he would add that this conception squares well with our commonsensical scheme for describing and explaining actions. Just how commonsensical it is is a question that must await discussion of what is made available to the interpreter in interpretation, and in his choice between

schemes. At the moment, however, I want to turn to the question of whether there *is* any way of relating the psychological scheme more closely to the underlying physical conditions, notwithstanding holism.

This is to raise the issue of reductionism. Holism and rationality bind the intentional scheme of the ordinary interpreter that gives us, it is claimed, our only understanding of the mind. However, the holism of the mental is not sufficient by itself to establish the irreducibility of psychology to physics. For physical concepts no less than psychological ones can be defined by their place in a holistic scheme of explanation. When we apply a physical predicate, as Davidson puts it, "each case tests a theory and depends on one" (Davidson1980, p221). There is holism in the physical realm too. But although holism of the mental does not amount to an irreducibility claim, it is a step in the chain of reasoning that takes us there. As Davidson himself say, "Clearly this holism of the mental realm is a clue both to the autonomy and to the anomalous character of the mental." (Davidson1980 p217) I shall now follow up his lead.

2.8) The Irreducibility of the Mental

Davidson's claim for the irreducibility of the intentional level rests upon a second and more subtle application of the principle of rationality. Rationality enters this time to guide the continuous adjustments and revisions to an interpretation as more information comes to light. It controls each stage of this revising process as a normative standard, pulling the patterns of attributed attitudes towards the ideal of rationality; leading the interpreter to alter his future decisions and to re-evaluate those he made in the past.

Despite the delicacy of considerations concerning this appeal to rationality, the argument which trades on them is quite straightforward. The mental concepts constitutive of the intentional level are governed by a principle of rationality. If no similar condition governs the deployment of physical concepts, then conceptual incorporation, on which the nomological reduction to the physical level of description depends, will be frustrated. Let us reconsider the appeal to rationality.

Chapter Two

Rationality is constitutive of the range of application of mental concepts, like belief, desire, intention and action. It requires that we must give reasons for people's beliefs and actions which explain the presence of those beliefs and the performance of those actions. Rationality is an a priori principle of psychology; together with other a priori principles of interpretation, it delineates the permissible structure of the psychological realm. Thus the psychological realm is at least partly constituted by a priori principles. But does this show that psychology is immune to the claims advanced by empirical science? Not as it stands. The mere a priori status of principles in psychology does not guarantee the irreducibility of the mental. For a number of a priori principles govern the application of physical concepts, too; as, for example, in the case of the concept of length where judgements of the form 'x is longer than y' must respect the transitivity of that relation. If we are going to argue for the irreducibility of the mental to the physical we need to know more about the differences in content between a priori claims in psychology and a priori claims in physics. Davidson regards the difference as no less great than that between philosophy and science; he concludes that the study of the mind is the province of the former and not the latter. According to Davidson, we must formulate our criteria of the mental in the vocabulary of the propositional attitudes and when we do this we shall see that we must treat "Psychology as Philosophy". This conclusion is Davidson's resting place; it serves to constitute a defence of bedrock. What we need to know is how Davidson reaches that conclusion.

The principles of interpretation which govern the psychological realm are neither empirical generalisations, nor purely logical laws. They are a priori principles which relate items within the psychological realm to one another and to the empirical conditions of a subject's behavioural and physical circumstances. If they operated within the psychological realm exclusively, they would provide no behavioural grounds for the attribution of mental notions. They would provide necessary and sufficient conditions on rationality defined over sets of beliefs and desire alone. And yet the strong intuition is that we must have some behavioural expectations of creatures with particular beliefs, since not just

Chapter Two

any sort of behaving system can be accredited a range of complex thoughts. The behavioural and physical conditions of a creature must place constraints on how we interpret it (and the range of interpretations it can be given), and it will turn out that these conditions have to be specified in the interpretive scheme itself. So the constitutive principles need to give empirical application to intentional notions in such a way as to respect the differences between subjects' physical and behaviour circumstances. Clearly, this can be achieved without infringing their a prioricity, but there is more than one way to do this. One route is that taken by the logical behaviourist who seeks to provide analytic definitions of psychological notions in terms of complex empirical conditions on behaving subjects. This route ensures an empirical application for mental terms by means of principles that require no empirical support themselves, but it also treats the sought after empirical conditions on behaviour as exhausting the content of mental terms. This is essentially a reductionist thesis. If this was the only way for a set of a priori principles to give the concepts they govern an empirical application, we would be forced towards reductionism. However, there is no strict requirement that a priori principles should lack empirical content themselves, and that they should be analytic rather than synthetic. The principle of rationality might apply to physical creatures only if they meet some empirical condition discovered by investigation. This whole issue of analytic versus synthetic turns on whether these constitutive principles do treat psychological concepts as exhausted by some complex empirical condition for their application, as in the dispositional analysis of fragility; or whether they merely make ineliminable reference to the empirical conditions of their application, as in Kant's definitions of geometrical (and arithmetical) terms. If the principles of interpretation are analytic they will give us necessary and sufficient conditions for ascribing psychological states to others, and hence something akin to logical behaviourism; whereas if the principles are synthetic they will merely provide necessary empirical conditions for psychological ascription, and a looser connection of supervenience between the mental and the behavioural (or physical). Let us now examine these issues.

Chapter Two

In characterising our understanding of mental terms, Davidson wants to insist on an observability assumption:

There are conceptual ties between the attitudes and behaviour which are sufficient, given enough information about actual and potential behaviour, to allow correct inference to the attitudes. (1982, p476)

Does observability entail behaviourism? This is far from clear. To begin with, we cannot deduce someone's state of mind from a single behavioural episode in isolation. The holism of the mental dictates that a belief will show up in behaviour only as mediated by further beliefs, desires, and other attitudes. To say that any piece of behaviour reveals a belief, or is an action, is to say something whose truth and content rests upon the existence of a structure of other beliefs and desires, themselves at the service of other projects. They will interact with other beliefs, with desires, and with intentional states of all kinds to give reasons for what someone thinks and how he acts. So behaviour can exemplify particular states of mind only given certain assumptions about other states of mind, present and evident elsewhere. Correlatively, the same piece of behaviour can manifest different and sometimes contradictory attitudes so long as we make corresponding adjustments to the background reasons. All of which shows that a very complex pattern of intentional attributions attends the ascription of a single thought. And given that each item in this hugely elaborate and interlocking system of propositional attitudes and actions must show up somewhere in behaviour, this means that "a very complex pattern of behaviour must be observed to justify the attribution of a single thought" and that "unless there is actually such a complex pattern of behaviour, there is no thought." (1982, p476). So the best we can hope for is a holistic fit between a total interpretation scheme and a creature with a considerably complex repertoire of behaviours, which would seem to threaten logical behaviourism and analyticity.

However, holism of this sort only rules out a dispositional analysis of individual mental states, and so, in effect only rules out the most *straightforward* type of behaviourist reduction; it does not by itself

guarantee psychological irreducibility. Could we not try to work out a more sophisticated scheme of reduction, for instance, functionalism?

Davidson resists this suggestion but his reasons are hard to find. At one point he suggests that as interpreters we could not observe the whole complex pattern of a person's behaviour, so instead we must look for "good reason to believe there is such a complex pattern of behaviour." (Ibid). But what would "good reason" amount to here? One thing it isn't is a prediction based on the physical facts to further physical facts about the way someone will behave. Whatever it is, it must at least draw upon an implicit conception of how someone must act so as to satisfy the constraints we put on the interpretation scheme that we enter at that point of attribution. Having good reason to suppose someone capable of a range of intentional behaviours incorporates thinking of them as rational:

To have a single propositional attitude is to have a largely correct logic, in the sense of having a pattern of beliefs that logically cohere. This is one reason why to have propositional attitudes is to be a rational creature. (Davidson 1982, p475-6)

Rationality is used initially to provide the ties for the holistic connections that pervade and make for someone's mental life. But Davidson argues for the irreducibility of the mental to (physical) behaviour by a second and more subtle application of the concept of rationality. His argument is that we can distinguish between the synthetic a priori principles that govern the physical realm and those that govern the mental, because unlike their physical cousins, the synthetic a priori principles of psychology are not laws. When the interpreter speaks of having "good reason" to attribute some mental state to a creature, his reason cannot be formulated explicitly as a statement of "what evidence is necessary or sufficient to determine the presence of a particular thought."

This involves Davidson's claim that "the way desire and belief work to cause the action [they explain] must meet further, and unspecified, conditions." Knowing these further conditions *would* enable us to predict someone's behaviour and thought. But Davidson thinks these further conditions *cannot* be specified and that prediction is always and in



Chapter Two

principle beyond us. There are three reasons for this. Firstly, we cannot isolate the physical cause that is the precise reason for action. Secondly, there are no psychological laws. And thirdly, there are no psychophysical laws. And if we cannot make lawful predictions of physical consequences from a psychological description of events, then we cannot fit psychology into the nomological net of physical theory. The intentional level remains irreducible.

Let us consider the three reasons and their inter-relations in this argument. Firstly, we have Davidson's claim that there is no physical-level means of expressing which reasons cause an action. This seems to be because reasons are dictated at the rationally explanatory level. This is a level at which one can know the reasons for action while knowing very little about the underlying causes of the behaviour. To find the reasons for someone's action is to find the cause: causes are the physical concomitants of the beliefs and desires that serve as reasons. But this is not yet to show that there is no way to isolate rational causes at the lower level; it is merely to claim that we can pick out the relevant causes without a scientific investigation of all the background conditions that could have caused change:

Explanation by reasons avoids coping with the complexity of causal factors by singling out one, something it is able to do by omitting to provide, within the theory, a clear test of when the antecedent conditions hold. (Davidson 1980, p233)

Although the singling out is done in intentional terms could we not find some basis for a causal prediction from the physical cause; or a prediction to it from other causal factors? It depends on what else must be brought into the picture. If the intentionally described antecedents line-up with neural or neurophysiological states of the system, then perhaps we can. But what are the causal antecedents to action? Again we begin from the intentional standpoint. When, in interpretation, we identify an action with a physical or behavioural event we must make sure that the causal history and circumstances of that event include every physical event or state identical with any of the psychological states invoked in the rational

Chapter Two

explanation of that action. (see Davidson1980 p254) But any intentional state mentioned in that explanation owes its identity to its place in a network of other propositional attitudes to which it is rationally tied. So perhaps the best we can say is that the whole of the agent's psychology, including that reason which is the cause of the action in question, describes events in the agent, characterisable in physical terms, that include the physical cause of his physical movements: a part of the total mental life is identical with a part of the physical whole. But we have no way of carving up these parts and wholes across the levels to make closer and more local links between the mental and physical. A part of the whole of a creature's mental life is at the same time a physical part of the creature that causes this event in his physical history. (This is essentially Jennifer Hornsby's conclusion in Hornsby 1980 and 1985). But can this satisfy Davidson? What he is left with is the thesis that a whole system behaving in these diverse ways is describable in terms of particular thoughts and reasons for action, where we can find no closer links between the physical causes and the reasons for particular actions. However, to say that there are always inner causes for the behaviour of a person who can be ascribed some explanatory reasons is too weak to ensure that such reasons are the explanatory reasons for that behaviour. According to Davidson, they count as the agent's reasons for performing the action only if they cause it *in the right way* :

The point is not just any causal connection between the rationalising attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. (Davidson1980 p78)

To explain an action in the right way the beliefs and desires must be token-identical with physical states that must cause it in the right way:

What I despair of spelling out is the way in which attitudes must cause actions if they are to rationalise the action (Ibid., p79)

A physical description of the causal process would be insufficient if it left out the notion of a *good reason* for so acting, and an intentional level

Chapter Two

description conforming to this explanatory ideal merely returns us to the notions whose role we set out to fix more precisely. How are the intentional specified reasons to direct our sight to the relevant causal process within the agent when it is causality that determines the explanatorily operative reasons? Davidson seems to have drawn a blank. And yet this was the place where we might have hoped for closer links between the mental and the physical domains, and some guarantee of the substantive nature of rational explanation. Davidson can only hint that:

In the analysis of action, mention of causality takes up some of the slack between analysis and science (Ibid., p80)

At best, it is rather convenient for the a priori theorist that the causal factors are so elusive; at worse to advert to causality is just wishful thinking.

It is largely because we cannot see how to complete the statement of the causal conditions of intentional action that we cannot tell whether, if we got them right, the result would be a piece of analysis or an empirical law for predicting behaviour (Ibid. p80)

For it to be possible to complete these conditions we must be able to enumerate the states that make up the subject's mental life at some time, listing everything in that panoply of feelings, urges, thoughts, and fears the person has. But this is not possible either. Here, we have the second reason for being unable to specify the further conditions: the absence of psychological laws.

What is needed in the case of action, if we are to predict on the basis of desires and beliefs, is a quantitative calculus that brings all relevant beliefs and desires into the picture. There is no hope of refining the simple pattern of explanation on the basis of reasons into such a calculus. (Davidson1980, p233-4)

There is no way to make the psychological realm more deterministic and more like the physical realm. There are no strict psychological laws on

Chapter Two

the basis of which we could make the reliable predictions; laws, that is, couched in terms of psychological vocabulary alone. One reason for this is because the psychological realm, unlike the physical realm is not a closed system.

There couldn't be a closed system of the mental, because of the endless ways in which the mental interacts with the physical. (Davidson, 1985 - Vermazen and Hintikka p249)

The empirical generalisations we would need to predict someone's mental life with any accuracy would have to include conditions which could not themselves be described as mental. These would include the environmental conditions of perception, and the thinker's behavioural effect on his environment. But this is ruled out by our third reason: there are no psychophysical laws. The psychological concepts employed in the formulation of such laws must answer to one set of standards and the physical concepts to another and wholly different set of standards. The important difference in the standards for applying mental concepts is not that the attitude or action ascribed belongs to a holistic pattern, the same holds for the physical concepts; the difference lies in the nature of the conditions placed on how the pattern is to be elaborated from the place at which we first enter it.

Each interpretation and attribution of attitude is a move within a holistic theory, a theory necessarily governed by concern for consistency and general coherence with the truth, and it is this that sets these theories forever apart from those that describe mindless objects, or describe objects as mindless. (Davidson1984, p154)

In short, the holistic structure of the mental is a normative structure. The principles of interpretation give us a theory of mind and these constitutive principles have normative force in the mental realm, telling us what someone who is attributed certain propositional attitudes *should rationally* think, and what someone who thinks like that *ought rationally* to do and say. This is the normative structure of reason itself.

Chapter Two

The difference between psychology and physics rests on the idea that the norm of rationality is constitutive of the range of belief, desire, intention and action, and is an ideal to which revisions in the patternings of these concepts must be continuously adjusted.

It is this normative ideal of rationality which sharply separates psychology and the physical sciences, as John McDowell usefully makes clear:

To recognise the ideal status of the constitutive concept is to appreciate that the concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be. This is to be contrasted with a style of explanation in which one makes things intelligible by representing their coming into being as a particular instance of how things generally tend to happen. (McDowell 1985 p389)

The point to grasp about explanations in accordance with an ideal is that where people are imperfect, the explanatory force of the ideal is in pulling developing explanations into conformity with this standard. The rational explanations of behaviour will be peculiar to an individual; they will remain open-ended, partial and *post-hoc* interpretations. Any empirical generalisation we try to coin as a result of success in interpreting what an individual says or does,

.....merely states part of what we mean, in this context, by saying an action is rational. The only empirical generalisation on which the explanation rests is given when we describe [that person's] beliefs and desires. (Davidson 1980, p267)

Davidson frames the non-predictive, responsive posture of the interpreter like this:

.....when we use the concepts of belief, desire, and the rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency: the constitutive ideal of rationality partly controls each phase in the

Chapter Two

evolution of what must be an evolving theory. (Davidson1980, p223)

He calls this evolving process of interpretation "opportunistic tampering". This lays important stress on the *ex post facto* element in the interpreter's decisions as to what thoughts or feelings to attribute to a person. The interpreter will look for revisions in the light of what best explains the subject's behaviour as more of it becomes known. This epistemological holism requires the interpreter to adjust each act of interpretation to all others.

There is a useful contrast here between Davidson and attributionists like Dennett who stress the predictive aspect of propositional attitude psychology. For Dennett, intentional system theory is a rationality-based intentional psychology with predictive laws rather like those of rational decision theory. But he argues that such a theory describes an ideal that is seldom (never) met. It serves as an heuristic for predicting behaviour in terms of the beliefs and desires attributed, but it cannot explain because our psychology is just not like that. So we do not have these well-regulated beliefs and desires. This distinguishes Dennett's instrumentalist stance on attributionism from Davidson's realist stance. Both believe that the concept of belief is just the concept of an attribution made to someone on the basis of his behaviour and in accordance with a priori standards. It is in the attributing (or attributability) of mental states that they come into being. (Contrast this highly metaphysical picture of the coming into being of a mental state, courtesy of interpretation, with, in McDowell's terminology, their coming into being as a matter of how things tend to happen.)

People have mental states like belief and desire, if rationality, or the a priori conditions require it. Dennett believes that in the world of human experience, the ideal rationality conditions are not met and so people do not have beliefs and desires. This is the source of his instrumentalism. Davidson, on the other hand, thinks rationality does apply to human beings and so people do have beliefs and desires. This amounts to Davidson's realism about the attitudes. Both Davidson and Dennett would agree that human beings are not ideally rational. So where do their

Chapter Two

differences lie? The answer is in their different understanding of the concept of rationality. Davidson thinks the conditions of rationality are fluid and flexible, in fact, much less rigid than Dennett's rational decision metric. For Dennett the conditions are so stringent they have no application to us. But with more flexible conditions, it is open to Davidson to see the mental notions which rationality governs as having serious application to us. The price Davidson pays for this flexible, attributionist realism is the indeterminacy of the mental, which I shall turn to in the next section.

Meanwhile, the irreducibility of the mental comes to this. There can be no nomological reduction because this would require the conceptual incorporation of psychological concepts within physical theory which is impossible if physical theory cannot accommodate the normative element of rationality. The fluid and normative notion of rationality has no echo in physical theory.

If one were convinced by these conclusions but committed to the naturalist project, one might think that the only hope for science would be to embrace Eliminativism of the Mental; the sort of position adopted by Stich and the Churchlands. It is a position which admits the irreducibility in principle but cheerfully gives up the vocabulary of the mental altogether. But this just isn't possible for us, as Davidson points out,

We cannot conceive a language without psychological terms or expressions - there would be no way to translate it into our own language. Of course there could be a part or fragment of the language that lacked psychological expressions, provided there was a (complete) language in which to incorporate or explain the fragment. (Davidson 1980, p244)

So if one believes that there could be a science of the mental, one has to resist both eliminativism and the Davidsonian position which serves as its premiss. I shall continue to lay out the Davidsonian position, for it is the one to repudiate if one holds out hope for a Cognitive Science with room for mental notions. What follows in the next stages of exposition reveals the shortcomings of the Davidsonian position. They are consequences that its advocates have never seen as a *reductio* of the

Chapter Two

position, and maybe that would be to put things too strongly. But I suspect that they are conclusions which depart both from common sense and from philosophical wisdom.

Chapter Three

Indeterminacy and The Mental

3.1 Introduction

As we have seen, according to Davidson, our interpretative descriptions of people's mental lives are not based on psychological laws governing their mental scheme, nor on psychophysical laws relating their psychological states to physical behaviour and the environment. Instead intentional descriptions are constructed, and revised, in accordance with certain norms of interpretation. These include the a priori principles of rationality, consistency and coherence governing the mental, together with other a priori principles linking mental descriptions to behaviour. A further set of principles will have to be given with specific application to linguistic interpretation.

These principles of interpretation leave room for a good deal of slack in giving mental characterisations. According to Davidson, there is no way to formulate explicitly the types of behavioural evidence necessary or sufficient for attributing a particular thought, depending as it does for its identity on so much else. Part of the trouble is that there is "no fixed list of beliefs on which any particular thought depends" (Davidson 1982, p475). But what is the force of these claims? Is it just a fact about human beings as interpreters that they cannot elaborate the whole panoply of thoughts and beliefs that identify a particular thought? Could it be that there are facts of the matter about the *interpretability* of an agent which fix the content of his thought and action, making it the case that he has one rather than another set of beliefs? Such a conception of facts about interpretability would be strictly compatible with realism about attitude attribution and would undermine any quick route to the thesis of indeterminacy of the mental from the paucity of evidence available to the ordinary interpreter. This is the correct line for a realist like Davidson to take, so we must look harder for his justification of the indeterminacy claims. They must be claims about indeterminacy in principle.

Chapter Three

3.2 Irreducibility to Indeterminacy

Let me just set aside another possible misunderstanding. There are places in Davidson's writings where he seems to say that because there is no single standard of rationality, and hence no unique way to capture someone's frame of mind, this makes room for minds which are not reducible to a complex of well-behaved physical facts. However, I think it would be hopeless to argue that the mental occupies the gaps left open by the fact that we can give no very precise behavioural specifications for application of the terms for mental description. This would leave us open to the charge that the terms picked out a loose set of behavioural equivalences. This is not Davidson's intention. Instead he carves out the autonomous realm of psychology and then points to an indeterminacy there. We have to have a prior understanding of the mental terms before we discover that their application to particular cases is indeterminate.

Some confirmation for this direction of argument is given in Davidson:

...concepts like those of meaning and belief are, in a fundamental way, not reducible to physical, neurological, or even behaviouristic concepts. This irreducibility is not due, however, to the indeterminacy of meaning or translation....It is rather the methods we must invoke in constructing theories of belief and meaning that ensures the irreducibility of the concepts essential to those theories. (Davidson1984, p154)

So I have chosen to present the argument in the order that Davidson acknowledges:

I did not.....argue from the indeterminism of psychological laws to the irreducibility of psychology, but from the irreducibility to an indeterminism in addition to the indeterminism of physics. (Vermazen & Hintikka, p248)

But what is the argument from irreducibility to indeterminism of the mental? The key lies in the fact that the methods for constructing theories of belief and meaning are all we have to go on in fixing the contents of

Chapter Three

someone's attitudes and utterances (autonomy); and these methods will not prove sufficient to render interpretation determinate up to the choice of a unique scheme of assignments (indeterminacy).

We have already found support for the first step of the argument, so we have to discover a way to get from that step to the claim for indeterminacy. Let us look for this connection.

Firstly, there are no reliable predictions to be made about the mental. Mental life depends quite extensively on non-mental conditions, so there can be no purely psychological laws. Moreover, there is no hope of taking these other factors into account for there are no psychophysical laws. And so,

we have the Principle of the Anomalism of the Mental: there are no strict laws at all on the basis of which we can predict and explain mental phenomena (Davidson 1980, p224)

Anomalism does not straightforwardly entail radical indeterminacy. For we should not be tempted to conclude that in the absence of predictions from physical to mental conditions that there are no worthwhile connections between the intentional and the physical levels. Even in the absence of laws, there is still the method of interpretation to provide connections - by means of a priori principles - between the mental and the physical. However, in the absence of laws, the principles for imposing interpretations on suitable stretches of behaviour provide the only way of effecting the transition between the levels. What we need to know is how determinate interpretation can be. For an ideally situated and informed interpreter will know all there is to know about another person's mind. And it matters none if few of us can ever be in such a position, the point is one about interpretability.

But remember that interpretability must still be defined in terms of what could be available in human judgement. It must not transcend the point of view of a potential subject of experience, but merely extend it. So the claim now is that everything that is necessary for making the transitions from one level to another - everything necessary for picking out the facts about minds and meanings - is made available to the

Chapter Three

interpreter on the basis of observing the behaviour of another. This is the claim needed to defend the a prioristic approach to mental and meaningful matters that provides the first step in the argument from autonomy to indeterminacy. The claim will be that it is in making sense of others that we give intentional patterns to their behaviour, and that all we need to know about either minds or meanings is available, in principle, in those patterns as imposed on people's behaviour. So in the idealised case, it is the interpreter who effects the transitions from the lower to the higher level, by bringing the primitive data under the principles of interpretation. This is a conscious application of high level intentional notions to the bare physical facts.

I shall go on to dispute this claim with particular focus on the case of linguistic comprehension, insisting that it is not the speaker as interpreter who effects these transitions but information processing devices in speakers and hearers that accomplish it. In listening to speech in a language we know, there is a transition from the physical sound to the sentence understood. This process describes the speech chain from sound to meaning. The information we receive is a continuous sound signal and from that we must recover the syntactic structure of a sentence and, partly on that basis, the meaning it conveys. These transitions are conducted by special purpose, linguistic processes that take us from one body of information to another, extracting at each stage, the crucial information about phonemes, morphemes, syntactic structures, by chunking, ordering and reducing the information passed up as the output of the level below before handing on a product for further processing at the level above. High level principles guide the extractions and relations between the levels constrain the nature of information represented at each level. Not everything on which this information process depends is available to the rational interpreter or observable in a speaker's behaviour; nor can the information processing be given a purely physical explanation. The interpreter would have say that the process of linguistic comprehension could be brought to light by rational reconstruction. But it will be a bone of contention, in what follows, whether everything that we need to describe is made available by an purely a priori investigation that neglects

Chapter Three

empirical research. I shall argue that no such pure inquiry can account for our mastery of a language and that we need a more substantive account of the empirical conditions that underlie the understanding and use of a language than the attributionist can provide. Speakers are sensitive to the essential linguistic features that in part define the particular language they speak, but possession of this information is neither a matter of a speaker's reasons for acting nor the physical causes he is subject to. There is a need here to postulate content-bearing states whose contents are not available for conscious reflection to the subject of experience. I shall suggest that these states have a bearing of what a speaker knows. This will mean a re-setting of the boundaries between the conceptual and the empirical domains, making scientific explanation in the cognitive sciences relevant to the questions posed by the philosopher. But I anticipate.

Returning to rational interpretation, we are faced with the following situation. The determinate physical world is captured in the nomological net of physical theory. The anomalous mental realm cannot be so captured because there is no reduction (no set of concepts in physical theory) which can preserve the constitutive ideal of rationality that governs the domain of the mental. Therefore physical conditions do not *determine* the intentional cast of someone's mind:

Even if someone knew the entire physical history of the world, and every mental event were [token-] identical with a physical, it would not follow that he could predict or explain a single mental event (so described, of course). (Davidson1980, p224)

The epistemic limits of interpreters are irrelevant here. No set of physical facts can uniquely fix the list of beliefs on which the correctness of any attribution depends. So the question is whether the normative principles of rational interpretation which govern our descriptions of other minds can do this. Do they themselves fully *determine* the character of someone's mind?

Davidson's answer is that they do not determine a *unique* interpretative scheme. It is wholly consistent that a given set of physical

Chapter Three

facts, together with the principles that guide interpretation should provide more than one correct mental description of an individual. And given that this is all there is to go on, there can be no further fact of the matter to choose between equally adequate rival schemes. The mental is introduced as a level of description of certain physical events, those descriptions being licensed by the principles of interpretation. There is nothing else to draw upon when considering the mental facts.

I shall consider how these alternative schemes arise presently, but first, a few words about the consequences of this outcome.

Mental states are forged from behaviour by the principles of interpretation: mental life is that complex pattern of attitudes and actions by which the interpreter organises a person's behaviour and physical circumstances in intentional terms. Therefore a creature enjoys mental life just in case the categories of its intentional organisation are those of an ideal interpreter's scheme of explanation. There can be no gap between the best interpretation of an individual's total verbal and non-verbal behaviour and the facts about his mental life.

The obvious corollary of this is that the best that can be done by way of refining an interpretation is the best we can expect by way of degree of discrimination within mental life. Any indeterminacy in interpretation will be indeterminacy in the mental realm.

Just as in the case of holism, we can transfer properties of interpretation to what is interpreted. This is legitimate, by the attributionist's lights, since there is no gap between them. According to the attributionist, it is wrong to conceive interpretation as a good method for getting at the facts of mental life. It is not an epistemological thesis but a constitutive one. Its principles do not answer to the reality of the mental realm, they constitute it. These constitutive principles tell us what minds are. So essential properties of interpretation will make for essential properties of minds.

A special problem now confronts this picture: what are the essential features of interpretation? Earlier we drew a distinction between the conditions under which particular interpreters operate and interpretability in general. Since the essential features of the mind answer to

Chapter Three

interpretability, we must beware of drawing general conclusions on the basis of the facts about particular interpretations. Facts about interpretability must be determined by reflection on the principles of interpretation theory: an explicit and idealised account of what the fully informed interpreter knows. But now a second danger confronts us. Not every aspect of the theory of interpretation will be an aspect of the interpreted mind. We must somehow distinguish between the aspects of an interpretation that determine facts of the matter and those which are merely artefacts of the theory that does the interpreting. For example we may be able to provide a theory of interpretation in English for the total life and conduct of a monolingual Frenchman. Nobody would want to say that our subject really thought in English just because the interpreter was an Englishman. This is a trivial example, but more testing cases will emerge when we turn to the linguistic aspect of interpreting a subject. Should every aspect of the formalisation of linguistic theory be seen as displaying a property of the language under study? Davidson thinks not, but as we shall see, the distinction between essential and arbitrary properties of interpretation theory will be a very difficult distinction to draw.

Part of the problem here is that there is just no clear separation between theory and evidence for the interpreter. The evidence for an attribution must itself be interpreted and so its description becomes part of the constitutive claim that the agent has particular attitudes. The interpreter cannot but transform the merely primitive data of an agent's bodily movements into identifications of behaviour as intentional, even though he cannot fix their precise character until he has made many more intentional identifications. Either his total interpretations work out, or they don't, but he can tell us no more about this. Furthermore, he believes that no one else can either.

This is the least satisfactory part of the attributionist's project of interpretation. For we can offer no non-interpretative reason why interpretations work where and when they do. Officially, there is nothing in the agent for a correct theory to latch on to, or to be keeping faith with. The conditions for correctness are all to be found within the

Chapter Three

interpretational scheme. There are no facts interpretation is 'tracking'. With these worries in mind, let us return to the claims of indeterminacy.

Why should we accept these conclusions? Why not suppose this to be a case of the underdetermination of theory by data? The Davidsonian response to this challenge should be to produce an a priori argument in the spirit of attributionism about the limit of intelligibility of mental notions and their applications to behavioural cases. In this way the indeterminacy claim would be a consequence of reflections on the nature of fit between intentional notions and all possible evidence for them. I find no such explicit argument in Davidson but I think we can help him to one. It would go as follows.

The physical and behavioural circumstances of an agent do not uniquely determine his intentional patterns of attitudes and actions. Nevertheless, psychological descriptions of attitudes and actions *supervene* on the physical and behavioural circumstances of the agent. Supervenience in this context is the claim that there can be no mental difference without a physical difference; so any differences between individuals in their mental states demands that there be some difference in their physical conditions. (The same point holds for differences in the mental make-up of an individual over time.) Now consider the case of creatures whose physical histories can bear more than one description in terms of an interpretive scheme, so that events in their lives can be described as the performing of particular intentional actions relative to one interpretive scheme, and the performing of other intentional actions relative to another scheme. If it is a fact about their mental lives that they can be described by one interpretative scheme or the other, and by supervenience, there can be no psychological difference without a physical difference, then there can be no difference between these competing interpretative schemes. Neither can claim to present the mental life of an agent more accurately than the other. Either will do equally well at capturing what is there to be described. This is what indeterminacy amounts to.

This conclusion might seem to rest on too behaviouristic a reading of Davidson: if two creatures share all the same behavioural dispositions,

they will enjoy the very same mental life. It might be objected that this Behaviourist Principle is not the correct interpretation of supervenience since the supervenience here might not be local supervenience; that is, the behavioural might not exhaust the physical. The physical differences that could distinguish between two creature's minds might include their relations to objects in their environments. But now the claim would have to be that these external facts which causally impinge on the creature's sense organs give them different thoughts about different objects and hence lead us to give different interpretations of their thinking. However, Davidson supposes that the degree of discrimination in someone's thinking depends on the possibilities of discrimination which that person's language admits of. So there must be some way to distinguish between the words on one creature's lips referring to one object, and the same words on the other creature's lips referring to a different object. But, as we shall soon see, Davidson advocates an instrumentalist thesis about reference, supposing it to be a theoretical relation without empirical content. So where singular reference drops out of the picture the chance of making such fine discrimination in thought has gone. There will be no change in a subject's patterns of action or assent to sentences when one objective referent is substituted for another. So the relations of thought or reference to perceptual objects or referents in the subject's environment cannot be the relevant physical differences with which to differentiate between rival interpretations. (But note that Davidson is now prepared to endorse the points made by Tyler Burge about the dependence of mental content on facts about the thinker's physical or social environment. If he still holds to his instrumentalism about reference, then this marks a place where I am unable to render his position consistent. See Davidson, 1987 and Burge 1982, 1987) The upshot of all this is that if interpretation is indeterminate then so, to that extent, is mental life.

I have shown how one can argue from the indeterminacy of interpretation to the indeterminacy of mental life. Now I shall go on to say why there is apt to be indeterminacy in interpretation. Let us consider this.

There are two places to look for an indeterminacy in interpretation: one is between schemes and the other is within a single scheme of

Chapter Three

interpretation. The first is an issue about the correct attribution of a set of propositional attitudes, the second is an issue about the possession of concepts.

Alternative, competing schemes of interpretation that are found equally adequate by all standards of interpretation and all the available evidence will make different attributions of attitudes to a single subject. Although, as we have seen, indeterminacy of this sort may blur the difference between these sets of propositional attitudes. So in terms of psychological reality of mental life, these schemes will be two different ways to specify the same intentional facts.

But there is also indeterminacy within any interpretation scheme itself. This follows from general facts about interpretation together with Davidson's thesis of the inextricability of belief and meaning: "In attributing beliefs we can make very fine distinctions, as fine as our language provides." (Davidson1980, p238) But if we adhere to the inscrutability of reference, then attributions of meaning, and hence thought, will be much less fine-grained than we took them to be. This sort of indeterminacy blurs distinctions within the domain of conceptual thinking. Concepts as constituents of the thoughts expressed by whole sentences become theoretical entities without empirical underpinnings. The argument for this is rather involved, but I shall anticipate my discussion of it by noting the fate of the words we often use to express concepts; a use which provides one way of displaying our possession of those concepts:

I suggest that words, meanings of words, reference and satisfaction are posits we need to implement a theory of truth. They serve this purpose without needing independent confirmation or empirical basis (Davidson1984, p223)

Now in so far as the conceptual articulation of a thought depends on the semantic roles of constituents of the sentence that expresses that thought, concepts - like words - will prove theoretically useful but will correspond to no empirical condition of the thinker. This latter way of expressing indeterminacy is extremely pernicious in Davidson's hands,

Chapter Three

because where there is no further refinement in our thought, and where there is no scheme-content distinction (as Davidson maintains there cannot be), there can be no gap between the world and our categories of thought. Hence, indeterminacy reigns in the physical world too. This brand of empiricism is very thin indeed.

Leaving this vast topic aside I shall now take the two degrees of indeterminacy in turn. The first deals with different attributions of action, belief and desire. The second depends on the indeterminacy of linguistic interpretation and will involve us in issues we shall meet when discussing the attributionist's handling of the theory of meaning.

3.3) Indeterminacy Between Interpretational Schemes

Indeterminacy between schemes leads us at first to think that within any such scheme a good deal of the interpreted subject's mental life is fixed. Relative to that one scheme of attributions we can say precisely what our subject is thinking, but the choice between schemes remains, within limits, somewhat arbitrary. This thought is sponsored by the model of measurement in the physical sciences. Within any arbitrary scheme of measurement a very tight set of connections holds. Thus the scale of fahrenheit has very exacting standards, just as the scale of centigrade does. But there is just no fact of the matter in the mean kinetic energy of molecules to decide between the fahrenheit or the centigrade scales of measurement: either is correct. Choice does not matter.

This is the sort of indeterminacy Davidson envisages between rival schemes of interpretation:

Some choices among alternative systems of interpretation have no empirical significance; there choices are like the choice between metres and miles to measure distance (Davidson, 1985 p 252)

Now there cannot be complete freedom in the kinds of interpretation schemes we apply to a person. There will be limits as to what sort of behaviour counts as one's performing an action of a certain type. For

Chapter Three

example, certain minimal behavioural requirements have to be met before we can call an event an act of striking someone in fury. Nevertheless, as we move into the interior of the rational mind these minimal requirements leave a great deal of room for manoeuvre. Here, we have to decide which reasons are operative in the agent. And since action is behaviour backed by mind, the purpose with which someone acted will individuate the action more precisely. An intentional action is individuated by the beliefs and desires it introduces. The content of the intentional description under which an action was performed must be rationally tied to the propositional objects of the beliefs and desires that gave the agent his reason for acting. If we describe someone as humiliating his colleague and suppose that it was his intention to do so, then we are saying, in effect, that this was what he most wanted to do, that there were no countervailing circumstances, or unresolved conflicting desires, and that he believed that by doing what he did or saying what he said, he would achieve the desired effect.

We say of someone who acted with a certain intention, that he acted for a reason and spell out his reason by citing judgements, wants and wishes. But notice that this sort of explanation can go beyond any formulation of motive the agent can give of himself. As Davidson says:

As observers we often describe the actions of others in ways that would not occur to them. This does not mean that the concept of intention has been left behind, however, for happenings cease to be actions only when there is no way of describing them in term of intention. (Davidson1980, p229)

From the interpreter's stance, someone's behaviour counts as an intentional action, if and only if, what he says and does can be described by the interpreter in intentional terms. These descriptions belong to that huge apparatus governed by the interpretational theory of mind. And the key point is this: in respect of someone's intention in acting there is no one correct description to be given. We can change the description under which someone's action was intentional so long as we make compensating adjustments in our (holistic) attributions to him. Take the

Chapter Three

example given above, we can describe the agent as trying (somewhat insensitively) to be funny, because he wanted to be and because he believed (falsely) that the comment about her writing would amuse her. Numerous such re-arrangements are possible. In fact, as many as we can invent, subject to the requirement that we make good narrative sense of the events in the agent's life, this in turn being subject to the constraints of rational coherence and consistency on how we interpret and imagine the person acting elsewhere. Obviously, a number of narrative possibilities suggest themselves.

If we could pin down some of a person's propositional attitudes more precisely then we should cut down on the number of available intentional descriptions we could give of him. But we have already seen that there are no necessary or sufficient conditions drafted in psychological laws or psychophysical laws for an agent's having a single attitude. Furthermore, the a priori relations between reasons and actions cannot come into operation until we identify someone's actions more precisely, and we cannot do this until we know more about the mental states that back them causally. We need to make initial conjectures and then bring those into conformity with the a priori principles. We are left with a certain amount of theoretical slack in these reconstructions of a person's intention in acting, and so the mind given to us via this theory is only interpretable up to indeterminacy of belief ascription.

I agree with Davidson here to the extent that *unless* we can find more discriminating conditions on what is for someone to have a particular belief, or mean something in particular by uttering a given sentence, then we are subject to this indeterminacy between competing schemes. Where I disagree with him is that I think there are more discriminating conditions. But first, let us turn to the second way indeterminacy arises for Davidson.

3.4) Indeterminacy Within An Interpretational Scheme

Here we are not dealing with the unresolvable choice of which mental states to attribute to a subject, we are dealing with indeterminacy in the materials we have for making those attributions. This second sub-species

Chapter Three

of the indeterminacy of interpretation will involve us in considerations of the linguistic resources for ascribing and identifying beliefs. The argument is based on the interdependence of language and the propositional attitudes, and it argues from the indeterminacy of linguistic interpretation to the indeterminacy of interpretation; this will serve as a second instance of a premiss for our *modus ponens* from the indeterminacy of interpretation to the indeterminacy of mental life as a whole. Once again, the attributionist move is made by treating properties of interpretation theory as properties of the mental: if there is indeterminacy in the descriptions the interpreter can give of the intentional contents of people's attitudes and utterances, then to that extent, there is indeterminacy in the contents of their attitudes and utterances. This is the correct conclusion for Davidson since "an event is mental only as described".

This conclusion may well be the right one to embrace if you believe that there are no further physical, neurophysiological or cognitive psychological facts pertinent to determining what a speaker means by a given utterance; and that one's theory of mind and language can dispense with a theory of cognitive development and language acquisition altogether. Given that view, where else could one look for connections between mind and meaning but in the domain of the propositional attitudes? As I said above, I shall propose a different strategy that does give prominence to the sub-personal psychological details in examining a speaker's linguistic abilities, but for the moment let us stay with the attributionist, who believes that because ordinarily we have knowledge of other people's minds and know what language they are speaking, we know what minds are and know what it is to be speaking that language. What matters to this picture, and to the conceptions we have, is what is made available to us ordinarily in interpretation.

In a number of places Davidson has claimed that a certain degree of complexity and sophistication in a creature's thought depends on its possession of a language. (1985 p 250-252, 1982, and 1984 *passim*). It is not easy to state what degree of sophistication Davidson has in mind here. Mostly, he expresses the point in general terms:

Chapter Three

We cannot hope in any case to cope with the full range and subtlety of psychological traits without taking account of language, for the finer distinctions among desires and beliefs, thoughts and fears, intentions and inferences, depend on the assumption of a cognitive structure as complex as that of language and cannot be understood apart from it. (Davidson 1980, p225)

But elsewhere, he does give more specific examples. (See Davidson, 1982) The properties of thought required include: the subject's being able to make a subjective-objective distinction; interpretation being sensitive to the failure of substitution of co-extensive terms *salva veritate* (strangely!); and applicability of the apparatus of quantification to refining questions about the ontology of the thinker's world-view. All of which, unsurprisingly, add up to the distinguishing marks of belief contexts.

Davidson offers further conditions on being a believer. To have beliefs is to have the concept of belief, which requires one to be able to apply that concept to others. This he thinks is necessary if a creature is to grasp the contrast between subjective and objective judgement. And since the concept of belief arises in the context of interpretation, believers must also be interpreters of others. Ultimately, there is no asymmetry in the roles of interpreted and interpreter for the case of human beings. We are self-interpreting animals who have to make sense of ourselves and others in rational terms. What matters on this story are the categories for making sense of mental life - only in virtue of making sense of ourselves and others do we make sense at all. It is here that language plays a key role, as I shall now explain.

For Davidson, the rational mind of an agent is open to public determination by the well informed interpreter. He interprets others on the basis of their linguistic and non-linguistic behaviour. Now, in asking what is made available in interpretation, we are asking questions about the linguistic abilities of both the interpreter and his subject.

In the first case, what is largely at stake is the degree of discrimination exercised by the interpreter of someone else's mental life. This amounts to a discussion of the possibilities of thought as mapped out by the possibilities of language. As Davidson says:

Chapter Three

We identify and discriminate between beliefs, desires, and the meanings of sentences by attaching propositions to them....In trying to understand you, I match up my own sentences or propositions with your utterances and attitudes. My sentences are related to one another by logic, inductive and deductive, and are hooked to the world in various ways. I cannot ignore these properties of my propositions in interpreting you, since these are the properties that individuate and identify propositions. But given the multitude of considerations, and the richness of the field of propositions, there will not be one best way for me to interpret you. (1985, p252)

These are linguistic facts about the interpreter's materials for making attributions, facts which will appear when he gives an interpretation of another. But what about the language of the subject he is interpreting? Well, this is precisely where the lack of asymmetry between subject and interpreter is relevant. If the subject is a believer, then he must also be an interpreter. And to be capable of interpreting others he must have the linguistic means to mark discriminations in their attitudes: the attitudes he ascribes to them. But of course the materials he has for making sense of them will be the linguistically expressed concepts in which he makes sense to himself: the linguistic means he has of making discriminations in thought and judgement. So through interpreting his language we may know his mind. But now remember that these points will apply to the interpreter giving an account of this subject's language. The linguistic resources *he* uses will reveal something of *his* mind, and so in developing a theory of linguistic interpretation for the subject he is also giving an account of the sense he makes himself. So an account of the linguistic meanings of sentences in which a speaker expresses his attitudes both interprets his mind and reveals the fine discriminations in thoughts available to the interpreter on which the possibility of interpretation depends. The project of linguistic interpretation might now seem to hold out a promise of some general conclusions about the nature of mental life. For the account of meaningful speech in terms of which the interpreter can *make sense* of the mind and language of others will show him to have the materials for making sense: that is, to be a creature with a rational mind. So any account of the meaningfulness of another's

Chapter Three

language will be illuminating of what it is to make sense simpliciter. To learn the language in which someone interprets others is to learn the medium in which things make sense to that person, and to understand that language is to know his mind. The sentences he uses in that clause to report the beliefs of others will be sentences in which he can express his own beliefs. So to gain insight into people's minds we need to interpret their language:

Indeed, our best route to the detailed identification of intentions and beliefs is by way of a theory of language behaviour. It makes no sense to suppose we can first intuit all of a person's intentions and beliefs and then get at what he means by what he says. rather we refine our theory of each in the light of the other. (Davidson1980, p258)

But the same point applies to the interpreter too:

the only access to the fine structure and individuation of beliefs is through the sentences speakers and interpreters of speakers use to express and describe beliefs. (Davidson 1986, p315)

Notice in the following quote how easily Davidson slips from points about the interpreter's language to points about the subject's speech, exploiting the lack of asymmetry:

All the distinctions available in our language are used in the attribution of belief (and desire and intention); this is perhaps obvious from the fact that we can attribute a belief by putting any declarative sentence after the words, "He believes that". There is every reason to hold, then, that establishing the correctness of an attribution of belief is no easier than interpreting a man's speech. But I think we can go further, and say that the problems are identical. Beliefs cannot be ascertained in general without command of a man's language; and we cannot master a man's language without knowing much of what he believes. (Davidson1980, p238)

This puts paid to the thought that if the subject expresses his own beliefs on occasion, then an interpretation of what he says might seem to

Chapter Three

provide a more direct route to belief-attribution than the one considered hitherto. Speech is an intentional act and so it brings in its train the usual host of beliefs, desires, intentions and actions we need to explain it. For this reason, too, interpretations of speech must fit into a holistic scheme of action-explanations where the ideals of rationality, coherence and consistency have constitutive force.

There are also related reasons, more specifically linguistic, why interpretation of a speaker's language is no more 'direct' a method than the overall project of interpreting him. Explanations of speech acts cannot be given in isolation:

Part of explaining such acts is interpreting them, in the sense of being able to say what the speaker's words expressed on an occasion of use - expressed in his language of course. We have a full grasp of what a man said when he uttered certain sounds only if we know his language, that is, are prepared to interpret a large number of things he might say.....To interpret a single speech act, therefore, we must have a grasp of the speaker's unrealized dispositions to perform other speech acts. (Davidson1980, p255)

And so:

We interpret a single speech act against the background of a theory of the speaker's language. Such a theory tells us (at least) the truth conditions of each of an infinite number of sentences the man might utter (Davidson1980, p256)

The qualifications on this project now recapitulate those we looked at for intentional psychology:

What is clear, however, is that such theory construction must be holistic: we cannot decide how to interpret a speaker's 'There's a whale' independently of how we interpret his 'There's a mammal', and words connected with these without end. We must interpret the whole pattern. (Davidson1980, p257)

Chapter Three

This theory of the speaker's language must be embedded within an overall project of rational interpretation of his behaviour; where the interpretations given to the speaker's utterances must chime in with the attributions of attitudes made to him in order to explain his behaviour *tout court*. To meet this requirement, the interpreter must take many of his subject's utterances to be sincere expressions of the subject's beliefs, and he must find evidence for the speaker's possession of those beliefs borne out elsewhere in his behaviour. Moreover, the explanations we can give of a subject's actions will provide a necessary background when explaining his linguistic actions; for here too we must be able to explain the subject's speech act as intelligible in the light of the beliefs and desires, that give him reason to assert. What we need is a way of extracting the meaning and other intentional notions from episodes of speech; and for this we need a theory that has application to every such episode.

Most illuminating of all, will be an account of how we can make sense of someone from scratch without knowing either his language or his beliefs. This is the imaginary case for which Davidson proposes the project of **radical interpretation**: a procedure by which one aims to recover the beliefs and meanings of a speaker starting with neither.

In order to interpret verbal behaviour, we must be able to tell when a speaker holds a sentence he speaks to be true. But sentences are held to be true partly because of what is believed, and partly because of what the speaker means by his words. The problem of interpretation therefore is the problem of abstracting simultaneously the roles of belief and meaning from the pattern of sentences to which a speaker subscribes over time. (Davidson 1980, p238)

To a large extent the isolable beliefs, the intentions a speaker is taken to be exercising in uttering, and the precise meanings we can attach to the particular sentences uttered, are theoretical constructs that vary with one another in a holistic structure of attributions. Such a scheme will comprise those interlocking patterns of meaning, belief, desire, intention and action that make best overall sense of someone's total life and conduct. All of

Chapter Three

which suggests that there are no such independent entities as the beliefs or meanings described by such a theory:

What is important is that if meaning and belief are interlocked as I have suggested, then the idea that each belief has a definite object, and the idea that each word and sentence has a definite meaning, cannot be invoked in describing the goal of a successful theory. For even if...there were no indeterminacy at all, entities such as meanings and objects of belief would be of no independent interest. We could...invent such entities with a clear conscience if we were sure there were no permissible variant theories. But if we knew this, we would know how to state our theories without mention of the objects. (Davidson 1984, p154)

But to say that the notions of individual sentence meaning and belief are contaminated by theory, because they are essentially identified by theoretical terms, is not yet to rob them of all empirical content. If the principles that govern the theory specify conditions for their application to behaviour under empirical conditions, then we can avoid the conclusion that they are purely theoretical notions with no empirical content. This is a very big issue which takes us back to the question raised above: which aspects of interpretation theory are features of the objects interpreted and which are mere artefacts of theory. Davidson starts out with the suggestion that the notions of reference, words and word meanings are theoretical, but as I shall argue he has no way of isolating his conclusions about these from spreading to conclusions about belief and sentence meaning, at which point common sense evaporates into undifferentiated wholes or is transformed into a more discriminating theory. Neither of these positions are entirely satisfactory, and Davidson has no retreat. What he himself wants is a half-way house where he can distinguish between those semantical notions which are theory-laden but have application to actual linguistic practice and those which are purely theoretical and do not. All of the relevant notions arise in the context of theory, so he needs some way of deciding between those which really apply to a speaker or his language and those which remain internal to the theory. What is not so clear is how the attributionist can be entitled to draw this distinction.

Chapter Three

Before looking in more detail at this issue, let us just remind ourselves of what Davidson has said about it. He regards many of the ordinary features of a language like words, and word meanings, reference and even sentence structure, as theoretical notions, articulated to implement the theory, having no empirical content of their own. Notice that this shows that Davidson believes there is still a question about what *is* to be found in the interpreted empirical evidence.

Given Davidson's denial of substance to the intra-sentential notions, it is as if, in the words of the poet Robert Frost, "A sentence is a sound in itself on which sounds, called words, may be strung". These whole sounds are the only evidential basis for fixing the empirical content of a theory of linguistic interpretation.

Note that Davidson thinks that neither the interpreter constructing a theory nor the incipient language user learning the language are ever "in a position directly to learn rules for assembling...[words].. into meaningful wholes." Instead the account should take the following broad outlines:

We start rather with the wholes, and infer (or contrive) an underlying structure. Meaning is the operative aspect of this structure. Since the structure is inferred, from the point of view anyway of what is needed and known for communication, we must view meaning itself as a theoretical construction. Like any construct, it is arbitrary except for the formal and empirical constraints we impose on it. In the case of meaning, the constraints cannot uniquely fix the theory of interpretation. (Davidson1980, p256-7)

Once we have fixed the formal and empirical constraints we shall have an adequate theory of meaning. And,

Guided by an adequate theory, we see how the actions and dispositions of speakers induce on the sentences of the language a semantic structure. (Davidson1984, p8)

But the question now is what can such a theory extract from the evidence, and more importantly as far as the attributionist is concerned, what can it read back into that evidence? Let us consider these as separate

Chapter Three

questions, pace the attributionist, and answer them in turn. I want to make a few remarks here about the former, and reserve the latter for the next chapter.

The answer to the first question is not very hopeful. We start with the cases of partially interpreted attitudes of a speaker: cases where, for some sentence *S*, the speaker **holds that sentence true**. This is somehow recovered from his behaviour without either knowing what he means by *S* or what he believes about the prevailing conditions for holding *S* true. We must suppose that this sentential attitude is manifested by assertion, assent, or dissent. But even to re-describe someone's linguistic behaviour to this extent is to have postulated two interlocking notions: what he means and what he believes. It is our attempt as interpreters to solve the simultaneous equations for these cases, subject to the formal and empirical constraints, that lead us to ascribe beliefs to the speaker and meanings to his words. Often we start with a defeasible assumption that our subject believes what we do about what is taking place when he assents to a given sentence. This is Davidson's appeal to the Principle of Charity. Now because I attribute to a speaker a belief of mine about what is the case, and I suppose myself to be knowledgeable about the prevailing conditions, I can suppose him to take the sentence we hold true to be true in these conditions. Thus my task as interpreter is to work out the truth conditions for this and all the other sentences he holds true, and from here try to project to all the other declarative sentences of his language. Since each ascription of meaning necessarily involves an attribution of belief, this linguistic project will interact with a theory of his beliefs. This becomes an asset in mature interpretation. Once we have fixed a meaning for his sentences - by a theory that enjoys best fit with the evidence of the sentences he holds true and all other ascribable attitudes - we can, on a future occasion, then we can interpret him as having a false belief about what is the case, if he assents to a sentence which is false. Belief takes up the slack between the interpreter's assignment of truth-value to a subject's sentence and his assent to, or dissent from, that sentence. However, if there is evidence to suggest that his belief about what transpires is veridical, then we have to interpret him otherwise, as having attached a

Chapter Three

different meaning to one of the sentences than that given to it by other speakers of his language. The point is that it is possible to choose to interpret him one way or the other, and even once all the evidence is in there may be no fact of the matter which allows us to choose between the two options. This is as expected:

...belief and meaning cannot be uniquely reconstructed from speech behaviour. The remaining indeterminacy should not be judged as a failure of interpretation, but rather as a logical consequence of the nature of theories of meaning (just as it is not a sign of some failure in our ability to measure temperature that the choice of an origin and a unit is arbitrary).
(Davidson 1980, p257)

I take it that it would be a failing, however, if we simply accepted that this is all there is to say about temperature. The increased energy in the molecules at constant pressure explains why we are likely to be measured as increasing in temperature. Perhaps in the case of meaning, too, there are facts at other levels of description to underpin our confidence that we were giving the right ascriptions?

Why should we see something like Davidson's theory of interpretation as the only way of approaching the nature of linguistic meaning and understanding? Could we not begin to draw on existing empirical research into natural language meaning and grammar to extract a more precise account of what a speaker means? After all, Davidson is seen here as saying that indeterminacy is a consequence of the theory of meaning he adopts. And as before, we conclude that what is available to the interpreter depends on the material and methods he can call upon.

Davidson's resistance to any such empirical, science-based alternative comes firstly from his insistence on what he thinks of as the common sense nature of his approach to mind and meaning, and secondly, by supporting this claim where he needs to by the metaphysical tenets of his attributionist position. It seems a plausible reading of Davidson to suggest that his reasons for maintaining that there is no other theoretical level of description and explanation to locate facts about meaning, in either the

Chapter Three

behavioural, brain, or cognitive sciences, are to be found in the philosophy of psychology already reviewed. We shall see how well they stand up later.

We can now see how Davidson's claim about the theoretical status of individual sentence meaning infects his theory of belief. Belief and meaning are interlocking notions: unknowns that must be solved for together since the choice of one informs the choice of the other. Now if the best way of reporting what someone believes is by a comprehending use of the sentence by means of which the subject expresses his thought, then any indeterminacy in our interpretations of his sentences must lead to that degree of indeterminacy in the thought we attribute to him; and, where we are best placed to interpret him, this means a corresponding indeterminacy in his thought. Where there is no means of saying precisely who X is speaking about, there is no fact of the matter concerning who X is thinking about. It is no good supposing that the speaker though not the interpreter might know who he was thinking of, so that there would still be a fact of the matter as to whom the speaker was referring to. For the only materials that can determine whom he is thinking of are those of belief-desire psychology in the rational theory of action, together with a theory of linguistic interpretation. There can be no more to the mind of a speaker than is available to a well informed interpreter, and the interpreter can only interpret a speaker's use of a singular term by specification of its reference in an interpretative truth-theory for that speaker's language. Any limitations on the empirical content of truth theories that serve as theories of linguistic interpretation will be limitations on the rationally interpreted mind itself. Any indeterminacy in interpretation is indeterminacy in thought. So it is now crucial to know how much indeterminacy infects meaning: just which properties of interpretation are properties of thought. I turn to this now in the light of Davidson's theory of meaning.

Chapter Four

Interpretation and Meaning

4.1) The Theory of Meaning: A Fregean Overview

Without the far-reaching and insightful work of Frege we should have very little idea of what a theory of meaning looks like. Although it was never fully worked out, Frege's model for the theory of meaning provided one of the clearest attempts to isolate the project that a finished theory of meaning would complete. Nearly all existing semantical frameworks owe something to Frege's original ideas. So perhaps it would be useful to provide a very quick sketch of this project before assessing Davidson's version.

Fregean theories of meaning are based on three key notions, derived from two key distinctions: force/sense/reference. Starting in the middle, the notion of sense is what we need to explain *what is said* by a speaker when he utters a sentence with assertoric, interrogative or imperative force. We see a distinction here between sense and force because Frege thought that the very same sense of an utterance, ie. what is said, can be what is asserted, what is commanded, what is asked on different occasions. So there can be sentences which express the same sense but where those sentences are uttered with different forces. The sense of a sentence (what is said) is not a property of the speaker of that sentence. Unlike an idea, different thinkers or speakers can grasp the very same sense on hearing a particular sentence uttered. To know what is said by uttering that sentence, whatever particular force attaches to the utterance, is to know something that is not the property of any one speaker's mind.

Sense is also distinct from reference. There is a difference between what is being talked *about* and what the speaker *says* about it. When we understand an expression we know more than just its reference; in particular we know the way it is picked out by that expression. However, in some cases we know much less than the reference, as in the case of informative identity statements. Frege takes it that which objects and properties we succeed in speaking about depends entirely on what is said; and so sense determines reference (the things talked about). The

sense/reference distinction is operative at many levels in the theory. At the word-level there is a distinction between the object referred to and the way it is referred to; there is also a distinction between the range of things a predicate applies to and the way of identifying the property they all have. At the sentence level there is a distinction between the truth or falsity of a sentence and what it typically asserts. Frege unified these different sense/reference distinctions in a general account of these two notions so that sense and reference became two distinct levels of meaning or content, with sense determining reference quite generally.

In the general account, Frege propounded the principle of compositionality. The sense and reference of complex expressions depend on the senses and references of their parts, respectively. But the dependencies are different in each case. At the level of sense, the overall sense of a sentence contains the senses of its subsentential expressions as constituents. While at the level of reference, the reference of a sentence - its truth-value - does not *contain* the references of its subsentential constituents. Instead the truth value is a *function* of the referents of those expressions. Both the selection of reference-determining constituents and the various functions from referents to truth-values depend on the semantic structure of sentences. However, it should not be thought that the semantic parts of a sentence could be given meaning (sense, or reference) in isolation. Their suitability in building up complex expressions was thought by Frege to be due to the semantic roles they enjoyed in molecular sentence structures: they are atoms that always occur bound in a molecule of meaning. This picture gave the sentence a certain primacy in discussions of linguistic content. This thought was first expressed by Frege in the preface to the Grundlagen in the form of a methodological maxim: never ask for the meaning of a word in isolation, but only in the context of a proposition. (Frege 1884, p x) Appearing as it does in the Grundlagen it is produced before the introduction of the sense/reference distinction. But it is evident that this primacy of the sentence underpins Frege's whole thinking about reference, and in the Grundlagen itself, it secures the role for his notion of reference in the argument for the existence of numbers as the abstract referents of

Chapter Four

numerals. The maxim has since come to be known as the Context Principle and any credible theory of meaning must comply with it. By means of it, we can think of the senses and references of words in terms of the stable contributions these recurrent items make to the senses and references of whole sentence in which they occur. The senses of words are suspended in the senses of whole sentences, and the reference of words can only be known through grasp of their senses:

...the determination of the truth-value of any given sentence goes via the identification of an object as the referent of each proper name occurring in the sentence: guided by the senses of these names, we first identify certain objects as their referents,.....[and then] decide whether or not the predicate or relational expression applies to that object or those objects. (Dummett 1973, p229)

What this account rules out is *bare* knowledge of a term's reference, where this would be a way of knowing the reference where there was nothing more that could be said about a speaker's ways of knowing it.

Finally, Frege used the notion of sentence sense to play the role of object for beliefs and other propositional attitudes. To characterise the content of a thinker's belief we need a notion of content more discriminating than just whether the belief is true or false, or just which things the belief concerns. The notion of reference will not provide the objects of belief. Thinkers do not just believe that a proposition is true or false, they believe what the proposition propounds. When we choose a sentence to report someone's belief we want to say that he believes *what is said* by that sentence, and sense is the notion we need to provide an account of *what is said*. This relation to belief enables us to provide a criterion for the senses of sentences based on the sameness or difference in cognitive value of two sentences. Two sentences will differ in cognitive value just in case it is possible for a speaker who understands both of them to take different attitudes towards them at the same time; believing one but not the other, say. This is possible even when the two sentences share exactly the same referents and structures.

Chapter Four

It is a moot point whether or not Davidson's theory of meaning is truly Fregean. This turns, finally, on whether he has any use for the notion of sense. Commentators like McDowell think he has, but as we shall see there are many reasons to doubt this.

4.2) Truth and Interpretation

Davidson opts for a truth theory as his theory of meaning. But since on the face of it this is a semantic theory that would correspond to a theory of reference in the Fregean framework, we might wonder how it can be capable of fully specifying the meaning, or sense, of expressions and sentences in the language. Clearly, we need some assurance that a theory of truth can *serve as* a theory of meaning.

The claim is made in some circles that carefully selected theories of reference can serve as theories of sense. The idea is that in stating a theory of reference we have to choose some means of specifying the reference of any given expression, and that the linguistic means used to state the reference will *show* the sense of that term. This idea is fine as far as it goes, but the interesting work resides in saying what it is for a chosen means of reference-specification for a language to express the senses expressed by speakers of that language. For this we need to know the factors that constrain the correct choice of sense-displaying reference clauses. Presumably, these will be cognitive constraints on a theory of reference that ensure that it is adequate to the task of capturing the meanings of words and sentences. It is a large project to say what these constraints are and how they are motivated. I want now to look at this question in the context of a Davidsonian theory of meaning which takes the form of a Tarski style recursive characterisation of truth for a language.

Let us begin by asking what is it for a theory of truth to serve as a theory of meaning for a natural language? Perhaps one could find some suitable adequacy constraints that a truth theory for a language must meet if its theorems are to present the meanings of declarative sentences in that language. What we need to know first is how a Davidsonian decides that a

(Tarski-style) truth theory is a plausible form for a meaning-specifying theory to take.

We can start with the intuitive thought that the truth of a declarative sentence is determined by its meaning and the relevant worldly facts, we can treat this as a function called Evaluation:

E(meaning, facts) \rightarrow truth.

Now for any two-place function there are exactly two inverse functions, which we can call here, Knowledge, K(meaning, truth) \rightarrow facts, and Interpretation, I(facts, truth) \rightarrow meaning. The first of these concerns the transmission of knowledge by communication, but it is the second that concerns us most. For if we can provide a theory that supplies patterns of assignments as arguments for the I-function it will also yield specifications of sentence-meanings. This must be a theory that coordinates assignments of truth with the facts that make those sentences true - or, if the theory is to be sufficiently sensitive to the speakers' perspective - coordinates assignments of truth with (an interpretation of) speakers' beliefs about the states of affairs in which those sentences are true. An overall interpretation theory that yields pairings for the I-function will include a recursive means of attaching the truth-predicate to sentences of the speakers' language, together with a theory of their beliefs about what is the case when those sentences are true. The argument pairings served up by combining a theory for applying the truth predicate with a theory of belief can then be tested against the evidence of the sentences that speakers' holds true; and the best combinations will be those which make them *hold* true more sentences that *are* true (by the interpreter's standards), thus maximising the number of true beliefs. In these cases, the interpreter's route to linguistic interpretation is straightforward. If, when the speaker holds a sentence true his beliefs are veridical, and so concern what is the case (according to the interpreter), then the interpreter can know what is the case when one of the speaker's sentences is true. By taking the sentence to be true when those facts obtain, he will take that sentence to be *about* the obtaining of those facts. These cases give the interpreter local instances of interpretation. But to construct a reliable theory of linguistic interpretation, one needs to find a principled

way to coordinate the assignments of truth to sentences and beliefs to speakers, in the straightforward cases, to ensure correct pairings of arguments to the I-function for all other cases. That is to say, it is essential to find the right way to project beyond the evidential base to pairings of truth and truth maker that specify the truth conditions that native speakers would themselves attach to these sentences. Only then can one claim to have discovered the actual I-function for their language. Extensive application of that function to sentences previously unencountered or unconsidered will be given in the form of pairings that say for each sentence S: S is true when and only when p.

As we shall see, the principled means of achieving this is going to consist in providing a truth theory that meets certain formal and empirical constraints; the formal constraints will concern the mechanics of how the truth predicate gets assigned to sentences and the empirical constraints will have to do with when it can be correctly assigned to them. The claim is that philosophical illumination of the notion of meaning is to be had by examining the nature of these constraints and their interaction. For these are the constraints that together turn a formal theory of truth into an empirically adequate theory of linguistic interpretation capable of producing interpretive assignments of the conditions for truth to every declarative sentence of the language. Let us now look at these constraints.

The empirical constraint will be the serviceability of a truth theory's pronouncements in the overall project of rational interpretation of the agent. The formal constraints will have to do with whether the truth theory has the means to provide a recursive characterisation in the style of Tarski of the truth predicate for the speaker's language, specifying the conditions under the truth predicate applies to any given sentence. (Non-indicatives will have to be related to their corresponding indicatives by transformations supplied in a theory of force.) John McDowell has subsumed these two types of condition under the title of the Propositional Attitude Constraint and the Constraint of Systematicity, respectively. I shall state them both and go on to examine whether the empirical and the formal constraints coincide, and if not how we can motivate the latter. I

start now with the first requirement: that a theory of truth should subserve a theory of intentional interpretation.

4.3) Empirical Constraints on Truth Theory

What empirical facts ensure that a truth theory can interpret a speaker's utterances? Davidson's answer seems to point to everything relevant to the project of the radical interpreter. But not everything available to the radical interpreter is about meaning, so we need to know how meaning is determined in the face of the available evidence. Now a central aspect of radical interpretation is the interdependence of belief and meaning, and ultimately, their inextricability. The inextricability thesis is extremely important when defining the empirical constraints on linguistic interpretations, and much of what follows is an attempt to provide further clarification of it. But it is also an extremely subtle and elusive thesis, with many possible interpretations, so before I begin, perhaps it would be useful to list some of the claims I take it to encompass: (i) the evidence for linguistic interpretation by itself offers us no way of separating belief and meaning; (ii) we need a theory to identify the distinct contributions of each; (iii) belief and meaning are interlocking, complementary notions related at the theoretical level; and (iv) prior to bringing a theory to bear on the evidence these notions have no application to it. It follows that the only evidence for linguistic meaning is intentional evidence, and the only access to it is via a theory of interpretation. This is an extremely important claim to make, so we must give some consideration to how it emerges.

Davidson tells us that the subject matter of radical interpretation is the behaviour of a speaker, or speakers, and its task is to tell us what their utterances mean. Now on this project there are no detectable facts which concern linguistic meaning alone: evidence of language use is also evidence of the beliefs, desires and intentions of the language user. However, prior to interpretation there is no way of separating out, one by one, the distinct contributions of belief, desire and meaning. It is by application of theory that we recognise events in linguistic behaviour as

the upshot of the meaning the speaker attaches to a sentence and his beliefs and intentions in uttering it. Now an account that provides us with these sorts of descriptions will involve both a theory of meaning and a theory of mind. But now the question must be why do we need anything more than a theory of meaning to extract the linguistic significance of someone's utterance?

Well, the first thing to say is that the bits of behaviour that we are interested in are the deliberate linguistic emissions of a speaker: we are not interested in trying to make sense of those cases where a speaker is merely making noises. Speech is a highly intentional activity: the speaker puts his mind behind what he says. So we need to be able to identify those intentional uses of language if we are to embark on a proper course of linguistic interpretation. This suggests at least a minimal sense in which a theory of psychological interpretation must *accompany* a theory of meaning; but as yet there is no suggestion of a closer connection. However, this will not do for Davidson who conceives a much stronger link between the two theories. It is a link that arises from the exigencies of radical interpretation.

To interpret someone's linguistic behaviour is to say what his words mean on an occasion of use. The radical interpreter proposes to accomplish this by stating a theory that enables us to derive particular interpretations of each of a speaker's utterances. What we need to know is how to tell whether the interpretations it provides are correct. Merely to come up with an interpretation is no guarantee of correctness. There must be some standards for when an interpretation is right.

Perhaps this could be achieved in the following way. Our grounds for assigning meaning to people's utterances are also grounds for attributing beliefs: if someone assents to a sentence and we interpret it, we thereby attribute to him a belief. So perhaps the linguistic interpretations yielded by the meaning theory can be checked against our assessment of the speaker's beliefs and intentions. To interpret him as saying that *p* when we can attribute to him the belief that *p* is to have some reason for supposing that the sentence he utters expresses that belief for him. This conjecture could be checked if we found evidence that he had other

Chapter Four

related beliefs, and suitable desires which gave him reason to express that belief at the time of utterance.

There is something right about this but we must not make the mistake of supposing that we can fix the meanings of his sentences by a *prior* identification of his beliefs and intentions. The only adequate interpretation of an agent's mind includes an interpretation of his speech. This is not just because an interpretation of a person's language is a characterisation of his mind, although it is always that; it is the stronger claim that we could have no detailed knowledge of people's beliefs and intentions without having interpreted a large part of their language. This is because language use is a rich source of evidence about the fine structure and discrimination amongst an agent's attitudes. Furthermore, there is no way to complete a psychological interpretation without interpreting the person's speech. Speech is a form of intentional behaviour and so interpretations of speech are interpretations of action: speech acts. As such they have to conform to holistic schemes in which we provide rational explanations of the agent's behaviour in terms of the beliefs and desires that give him reason to act. So, given Davidson's view about the holistic character and scope of the mental, speech interpretation will be integral to the whole system of beliefs, desires and reasons with which we explain someone's behaviour; these interpretational schemes must include a way of specifying what people say, along with attributions of beliefs and desires that relate them to sentences and utterances respectively. So to make sense of someone's mental life demands nothing short of constructing a total theory for interpreting that person's speech, attitudes and actions on the basis of his behaviour: the mental scheme that emerges already includes the meanings of a speaker's words.

Thus far we have failed to find a way in which the radical interpreter can enter the circle of linguistic and intentional notions. There is no way to get at a person's beliefs and intentions without interpreting his language, and there is no way of arriving at a correct account of his linguistic behaviour without having both the linguistic and the psychological details right. The radical interpreter's task seems a daunting one:

Chapter Four

Since we cannot hope to interpret linguistic activity without knowing what a speaker believes, and cannot found a theory of what he means on a prior discovery of his beliefs and intentions, I conclude that in interpreting utterances from scratch - in radical interpretation - we must somehow deliver simultaneously a theory of belief and a theory of meaning. (Davidson 1984, p144)

The two theories are needed to pick out the two separate strands of interpretation in the partially interpreted intentional attitude of holding a sentence true. On any occasion in which the speaker takes up this complex attitude toward a sentence, the interpreter can treat it as having two interlocking components neither of which he knows in advance: what the speaker means by the sentence and what he believes about the conditions for holding it true. Strictly speaking, Davidson talks about the meaning *the sentence* has, and not the what *the speaker* means by it: for the Davidsonian interpreter is not in the business of defining the speaker-meaning of a sentence S in terms of a complex of psychological intentions, nor is he trying to identify a psychological state that amounts of a speaker's knowledge of the meaning of S; he is simply providing an account of the grounds for attributing beliefs to speakers and meanings to their utterances. It is a theory of the interpreter's competence, not the speaker's: it describes what every *interpreter* knows when he knows what is said; it does not describe how the speaker understands what he has said. Although, to succeed in capturing what can be said by any sentences of a speaker's language is *ipso facto* to provide an account of what the speaker knows in virtue of being able to speak the language. The point is that for the attributionist there can be no gap here. To interpret someone's language correctly is also to portray his psychological attitudes. There can be no further question of whether he really knows that *this* is what his sentences mean.

From the perspective of the radical interpreter, who can identify the sentences a speaker holds true before interpreting them, belief and meaning are just like the dependent variables of a theoretical structure: he must call upon some combination of theories of meaning and belief to

Chapter Four

provide plausible pairings of assignments so that he can try them out in interpretations of his subject's speech. The crucial aspect of this approach is that there is no way to fix one independently of the other. Davidson even goes so far as to recommend that "we should think of meanings and beliefs as interrelated constructs of a single theory" (Ibid., p146); a single theory which the interpreter has to bring to bear in interpreting his subject's behaviour, verbal and otherwise.

Intuitively, then, good total theories will interpret the sentences a speaker holds true as saying that *p* when they ascribe him the belief that *p*. And since not just any attribution of belief will satisfy the wider conditions on attitude ascription, the possibilities of belief-ascription will serve as a constraint on ascriptions of content to a speaker's words. But how can this constraint help us with our original inquiry, if it does not tell us how we are to arrive at good theories? It merely recommends that an interpretation should make someone say what we think he believes. But we started off with the question of how to tell whether a theory's interpretations of someone's linguistic behaviour are correct. All we have discovered is that the linguistic and the psychological details are complementary; that neither can be tested for independently of the other, and that a correct account must get both sets of details right.

Without committing ourselves to the claim that we can determine a unique set of pairings, we must still be able to judge whether any given set of belief and meaning assignments correctly interprets a speaker's sentences. Indeterminacy raises the further question of whether there really is a unique way to resolve these complex sentential attitudes into their two component factors. Davidson thinks there isn't: we could always change meaning assignments by altering someone's beliefs, and make adjustments to his beliefs by changing the meanings assigned to his words consistently with the overall constraints of rationality and holistic fit with the evidence. But these revisions can only be made within interpretation theory itself once we have a number of correct interpretations to consider. This is the force of the point that attributions of belief and meaning belong to a theoretical structure and are not independently established in the evidence; given the holistic nature of fit, there may be more than one

Chapter Four

way to give these two concepts their joint applications to the behavioural evidence.

All of this means that the empirical conditions on a correct meaning theory for a language are no more or less discriminating than the constraints on the overall project for interpreting its speakers as rational agents. A truth theory will serve as a theory of meaning when and only when its pronouncements meet this global test of interpretation. And the evidence for attributing a particular language to a speaker - as characterised by the truth theory - will be the very same as the evidence for attributing to him an overall pattern of beliefs, desires, intentions, actions and meanings. Since there is no merely empirical way of determining the character of that evidence, there can be no direct empirical route to determining what language a person speaks, or what he means on a given occasion, short of constructing an interpretation of his mental life:

For the evidence on which all these matters depends gives us no way of separating out the contributions of thought, action, desire and meaning one by one. Total theories are what we must construct and many theories will do equally well. (Davidson 1984, p241)

This means that finding a correct truth theory for the language of a group of speakers demands nothing less than finding an overall interpretive scheme for rationalising their behaviour; the evidence for an interpretational theory of truth - a theory which serves as a theory of meaning - will be nothing less than the evidence for a total scheme of interpretation that includes it. So how can Davidson claim that construction of such a theory will help the radical interpreter to arrive at a satisfactory total scheme for interpreting a person, if what determines the correct way of choosing a truth theory depends on success in interpreting all of a person's linguistic and non-linguistic behaviour? The answer he will give is that a theory of truth and a theory of the attitudes mutually constrain the means for achieving the overall task of interpreting someone. The radical interpreter must try out combinations of truth

conditions and attitude ascriptions to determine which combinations work best to interpret the total speech and conduct of the agent. A theory of truth from which one can derive truth conditions which pull their weight in this larger project will be a truth theory capable of specifying sentence meanings: that is, it will serve as a theory of meaning for that person's language. And now we have our standard of correctness for linguistic interpretation in the form of an empirical constraint on truth theories.

There are several points to notice about this empirical constraint. Firstly, it arises from theoretical decisions which can only be taken once we have entered the intentional realm. And secondly, it rules out any way of identifying a sentence's truth conditions at anything other than the intentional level. The theoretical framework relates meaning to the whole panoply of intentional attitudes and actions. Evidence for these cannot be reduced to evidence of a wholly non-intentional kind. And if meaning cannot be specified non-intentionally, then it cannot be located at the lower level. Therefore, meaning, like mind, is an irreducibly intentional notion, but one that cannot be reduced to the other irreducible mental notions. Admittedly, it follows from inextricability and the holistic nature of the interpreted mind that the most *direct* evidence for linguistic meaning the radical interpreter can hope for is the theory-dependent evidence of meaning *and* belief. But once he has entered the circle of linguistic and psychological notions, he can turn the inextricability thesis to his advantage to come up with a psychological constraint on any putative assignment of meaning. And so long as there are formal and empirical constraints of adequacy on the correctness of a theory for linguistic interpretation, there will be room to carve out a space within the interpreter's scheme for the notion of linguistic meaning. The empirical constraint on meaning assignments will be a theoretically motivated condition on the joint empirical application of the notions of belief and meaning. This is the condition John McDowell calls the **Propositional Attitude Constraint (PAC)**:

[A physically described piece of linguistic behaviour can be re-described by a theory of meaning as a saying with a specific

Chapter Four

content iff the re-descriptions are intelligible] against a background of propositional attitudes - centrally beliefs and desires - in terms of which the behaviour seems to make sense. Ascription of propositional attitudes, in turn, is constrained in complex ways by the physical facts about behaviour, the environment, and their interconnections; also (circling back) by the possibilities of interpreting linguistic behaviour in conformity with [certain formal requirements on the theory]. (McDowell, 1978, p122)

This *indirect* evidence gives us our condition on what it is for a formally adequate truth theory (I turn to the formal requirements shortly.) to be adequate for interpretation. Once the interpreter has opened up a space to begin interpretation, it tells us what success would amount to. An interpreter mindful of the empirical constraints must try to provide a truth theory whose content-specifications meet the PAC. These will be assignments of truth conditions that "pull their weight in making sense of speakers of the language we are concerned with" (McDowell 1977, p161)

Of course, by itself, this constraint does not tell us how a truth theory can meet such a requirement (given that we know something by now about how the constraints on propositional attitude ascription are to be met). It does not even tell us what it is for an act of utterance to be the expression of one of the speaker's attitudes: it merely insists that in specifying the meaning of someone's utterances we must come up with reasons for him saying these things. This may be part of the deliberate suggestion that although there is a close link between linguistics and psychology, there is no chance of a reduction of the former to the latter.

However Davidson does have a strategy for using the Propositional Attitude Constraint. His starting place is the sentences held true, and the starting assumption from the general theory of interpretation is that most of a speaker's beliefs are, by the interpreter's lights, true. This is the Principle of Charity. Now on that basis, Davidson hopes to hold belief constant and solve for meaning. By assuming the speaker's beliefs are veridical and are caused by the objects which prompt him to assent to a sentence he holds true, the interpreter can use his own beliefs about the

current situation to ascribe a belief to the speaker, and on that basis, a content to the sentence he holds true: it will be the content of the belief ascribable to both subject and interpreter that allows for the fixing of meaning. Notice how this view is reminiscent of Frege's identification of the sense of a sentence with a content of a belief. But the view is somewhat different. Frege arrives at this assumption from general reflections on content and the semantic evaluations of belief-sentences. Davidson arrives at his identification as a consequence of his methodology. For Davidson, the content of a belief coincides with the content of an utterance only when the objects that cause a particular belief also prompt a particular utterance. This is very different from Frege's idea of a shared cognitive content.

I said above that the principle of charity that maximises the number of beliefs which are true would most likely have to be replaced by a principle of humanity that tries to make the agent largely intelligible as opposed to just largely right. But if we take this step (and it seems the Davidsonian should), the strategy of holding belief constant and solving for meaning becomes that much less central. In fact, the grounds for a speaker's holding a sentence true would have to be fixed in the light of all the other attributions and so they themselves might be open to revision in the course of an interpretation. These would become less central also. But notice that these reflections do nothing to disrupt the force of the PAC. Combined with a principle of humanity, the empirical constraints on truth theory would come to this. A theory of truth adequate for interpretation will be one whose ascriptions of truth conditions together with a plausible theory of the attitudes makes best overall sense of a person's linguistic and non-linguistic behaviour.

4.4) Psychological Constraints on Truth Theory

We have seen how the *a priori* principles of interpretation can give rise to a **psychological constraint** on meaning, but it is a constraint governing the solution to a *simultaneous* equation involving the two unknowns. There is no way to use our *prior* knowledge of psychological

facts to direct us to the meaning of someone's words. The PAC simply repudiates any such epistemological asymmetry between psychology and semantics. We try out total patterns of belief and meaning assignment which we check against a speaker's behavioural evidence and the pattern of truth-values distributed over his sentences. This is a constraint on the successful completion of a total theory.

For there to be an asymmetry between semantics and psychology there would have to be psychological facts which the radical interpreter could appeal to in determining the meaning of hitherto uninterpreted linguistic behaviour; psychological details, that is, which were more empirically accessible than the deliverances of the theories of meaning they were supposed to constrain. (see Evans & McDowell 1976, p xv.) But of course that is just what is being denied. Empirically there is no way of getting at one before the other; the interpreter constructs his theory of meaning and theory of beliefs simultaneously, and tailors the output of one account to the other.

Certainly, this seems the right thing to say if one is considering *propositional attitude* psychology only, and if one is concerned with the evidence available to the radical interpreter. But one wants to ask whether there could be other ways of construing the psychological constraints on meaning. As I said at the beginning of this thesis, all parties to the dispute would accept the connection between mind and meaning and so would accept the existence of *some* psychological constraints on semantics, but the heart of the dispute lies with the precise nature of those constraints. To invoke a different type of psychological constraint than that favoured by the radical interpreter, we have to invoke a different conception of psychology. So, for instance, if we were concerned with cognitive psychology, there would be no need to think of the relevant psychological details of language users as being available to the radical interpreter - at least not consciously available to him as a subject of experience. The empirical facts of the sort documented in psychology of this sort may require elaborate investigation and detection, fitting as they do into a project of scientific explanation. Now some of these facts about speakers may be determined independently of the construction of theories of

Chapter Four

meaning for their language, and to that extent they may well be more empirically accessible than the semantical facts. However, it is just because those findings are not available on the basis of ordinary common sense reasoning that they are not considered by the Davidsonian to be either relevant or genuinely psychological: the right to deny an asymmetry between the linguistic and the psychological depends on a preference for rational a priori psychology and the linguistic predicament of the radical interpreter. But the rapid developments in the linguistic and cognitive sciences have given us reason to think that not all of the relevant psychological material is to be had at that level by ordinary reflections on observable behaviour. Facts about speaker's sensitivity to the constituent structure of their sentences may be unreportable by the speakers' themselves and yet these facts can be empirically confirmed and accessed independently of radical interpretation. Doubtless, the attributionist will deny the relevance of such so-called psychological facts to accounts of meaning. Indeed, he will take the findings of empirical cognitive science to have little to with the mental at all. This is due to his conception of intentional psychology and its claim for exclusivity: the only genuine psychological facts are the facts of rational psychology. So now we come up against two questions to ask of interpretation theory : (i) Is this the right description of our psychology? ; (ii) Do we need to consider facts about a speaker's language other than those available to the radical interpreter on the basis of ordinary reflection?

An answer to the first question depends both on comparisons with alternative conceptions of the mental, and on an answer to the second question. For if there are alternatives in the running, we need to know why we should choose one over another, and the answer to (ii) will provide the motivation. This will unfold slowly, so let me make a few points now in advance.

Where there are alternatives to his picture (though the attributionist will deny this), it will be no good asking him to consider which conception of the mind gets the facts about meaning and mind right, since he believes that the facts are determined by the theory, and the standards of correctness in describing those facts are internal to the theory. One

Chapter Four

wants to know how many of the ordinary intuitive facts about a language the radical interpreter can save in his account of linguistic meaning, but the interpreter qua attributionist works to slacken our grip on that way of looking at things, because he will insist that we have to collapse the distinction between evidential and constitutive claims. Our scheme gives our only grip on content, and the notion of pre-theoretic facts is put in doubt. It is interpretation theory which organises the welter of hitherto undifferentiated behaviour into descriptions of the mental and the meaningful. There are no prior linguistic or psychological facts for the interpreter to conform too because success in interpretation constitutes such facts. So the attributions of belief and meaning to an agent are seen through an interpreter's eyes; and what an interpreter singles out is something that owes its identity to its place in a surrounding theory. Each intentional description that can be given of a person's beliefs, or words, or actions supports many other attributions that stand or fall together. And now the gap between theory and evidence begins to close because we have to describe bits of behavioural evidence in terms that are themselves bits of the intentional theory. It seems that there is no getting at the *pieces* of relevant behaviour independently of the high level descriptions which organise and partition it. Of course, no single piece of behaviour is guaranteed its particular intentional description - these might be revised in the course of interpretation - but we always have to use the intentional vocabulary of the attitudes to describe behaviour in the first place, or we could not begin to filter out the sorts of linguistic activity the radical interpreter is interested in: the cases of speakers' holding sentences true. Now the vocabulary of intentional mental life demands a holistic fit with its behavioural base, and it is just this holistic fit which obscures the link between the theory of linguistic interpretation and its empirical evidence: it is the holistic nature of the fit that confounds the attempt to essay more local links between theoretical pronouncements and uninterpreted bits of behaviour. We do not seek out evidence for an interpretation theory: we try out theoretical descriptions as interpretations of the evidence in total.

Now there is only one way I can see to challenge this position directly - indirect challenges would come from eliminativists or others who seek to

Chapter Four

replace interpretation theory and its constructs wholesale - and that is to find instances of claims the interpreter is willing to make where he lacks the materials to make them. This would show that there was something seriously deficient with his picture. I intend to show that this is the case with the claims he wants to make about language. This will involve showing that there are certain claims about linguistic structure that are simply not available when we restrict ourselves to the predicament of the radical interpreter but which are assumed to be unproblematic. We might be tempted to think that the predicament of the radical interpreter is an epistemologically impoverished one and that it is only to be adopted when we are dealing with a wholly foreign tongue. Surely, we should not draw conclusions about all the facts there are about a language from the results of radical interpretation. For surely the radical interpreter's project is really unnecessary for discovering the empirical conditions on meaning assignment when we attempt to provide a theory of meaning for the language of speakers with whom we are able to converse. Here it seems the evidence is more local and the scrupulous methods of the radical interpreter are otiose. But this is not how Davidson sees it:

The problem of interpretation is domestic as well as foreign: it surfaces for speakers of the same language in the form of the question, how can it be determined that the language is the same? Speakers of the same language can go on the assumption that for them the same expressions are to be interpreted in the same way, but this does not indicate what justifies the assumption. All understanding of the speech of another involves radical interpretation. (Davidson 1984, p125)

Given the extensive application of this method to questions of linguistic interpretation we shall need to ask a supplementary question. Is radical interpretation sufficiently discriminating to bring the facts of linguistic meaning to light? In particular, is the methodology able to substantiate the claim that the radical interpreter can describe all of the linguistically meaningful sentences of a natural language. I shall suggest that some aspects of this claim are simply not available on the interpretive strategy. An argument of this form requires one to show that

the radical interpreter lacks the materials to describe what he claims he can describe quite explicitly. This is not an easy conclusion to prove since it requires a fairly exhaustive investigation of the resources available to the radical interpreter to make sure that he has no comeback to the challenge levelled at him. Hence my fastidious discussion of the Davidson programme. There is still a good deal further to go in examining the materials and the justification for the interpreter's claims to knowledge, for since the Davidsonian interpreter is concerned to answer a Kantian question of how knowledge of others' language and mind is possible we must not overlook any of the details that go into the answer nor include anything which is not detailed in the account. Davidson is offering not just a sufficient condition for such claims to knowledge, but also a claim that this is the only sufficient condition. In effect, what he is offering us is a transcendental argument telling us how meaning and mind (or our knowledge of them) is possible.

Now I want to claim that in giving an account of linguistic comprehension we do need to call on facts about language users which are not available on the basis of a priori theorising; not least because the concept of a language is not exhausted by a priori considerations. The detailed discussion of this issue will begin with the case of syntax, which I consider in the second half of this thesis. But for the moment I just want to prime the reader to this issue by drawing attention to the following point about the Davidson program. The cases of the sentential attitudes of holding true are pivotal in the project of radical interpretation: they can be identified without knowing either the meaning of the sentence or what the speaker believes about the world. They provide the point of entry into the speaker's intentional life, but at first these attitudes are identified by their external characteristics alone (cf. knowing that a sentence is true without knowing what proposition it expresses.) Initial descriptions of holdings true open up a space in the mind of the speaker to be further resolved into configurations of belief and meaning. This means that holdings true must be recognised as genuinely mental attitudes - attributions which carry commitments for the description of the rest of a subject's mental life. But they are also psychological attitudes that reveal

Chapter Four

intentional relations between speakers and bits of their language. And while the interpreter can know that a speaker holds a sentence true without knowing its meaning or the belief an utterance of it expresses, it is certainly not an attitude detectable without knowing which *sentence* the speaker is holding true. Now the term 'sentence' has both an everyday use and a more technical sense. It is the same notion that can be recognised by individuals in an ordinary way, and described by linguists in more technical terms. Ordinary speakers know which word strings are sentences of their language, but they do not know how they know this. Nor do they necessarily know what a sentence is. It is only when we turn to linguistic theory that we get an adequate conception of this notion. As linguists know, a sentence is not just an ordered string of words (as ambiguity shows); it is a hierarchically structured grouping of constituents which include both phrasal and lexical categories, and may include **null elements** which have no phonetic realisation. (eg. Verb-Phrase ellipsis in "Martin loves his cat and so does David -----.") These are intrinsic features that are not exhibited in anything like the surface structure of a sentence. They are described in syntactic theory and motivated by the gap between what the word string makes available and what is needed to determine the unique properties of a sentence with grammatical and semantical relations to other sentences of that language. Now since neither the a priori reflections of the interpreter, nor the facts about the physically realised utterance bring these details to light, we should conclude that our ordinary identification of sentences is neither a matter of reasons nor causes alone; the identity conditions of sentences are neither given by physics nor the folk, but by workers in empirical linguistics, mindful of the fact that ordinary people do not identify sentences in the terms linguists use to describe them. This makes it a very good question how the radical interpreter knows which sentences the speaker holds true. When the interpreter tries to work out the sentences of a foreign language he needs nothing short of a detailed theory of syntax which must be tested empirically and depends on facts about the speaker's linguistic cognition. In a language that the interpreter knows, where he is trying to determine whether he and the speaker mean the same by those

sentences, he can already recognise the sentences involved but not as a result of giving theoretical descriptions. There is a problem here which has been too quickly glossed over.

Detailed arguments for the existence of such syntactic features and their grasp by ordinary speakers must await my discussion of the work of Noam Chomsky, but let me suggest a controversial conclusion that will be made out in full in what follows. It is that when the radical interpreter postulates intentional relations between speakers and their sentences he is not simply alluding to facts about the existence of mental attitudes to be resolved by interpretation: he has already entered the mind of the speaker via the identity of the sentence recognised, for syntactic knowledge depends on certain internal configurations in the speaker. Identifications of the sentences a speaker holds true carry implications for the cognitive structuring and processes of his mind. And so far as he can identify them, they carry similar implications for the cognitive organisation of the interpreter. The full range of relevant syntactic details are neither conceptualised by the speaker nor the interpreter, they are neither physical processes nor propositional attitudes, but they are encoded in the cognitive states of a language user and put to use in processes that facilitate the necessary transitions from sound to meaning.

Now for Davidson it was the a priori theory which was supposed to help us to make these transitions from sound to meanings; and while I do not doubt that these transitions are made possible on the basis of everything available to the interpreter, they are not available to him as Davidson conceives it. It is not the radical interpreter, as a priori theorist, who brings all the linguistic facts to light, it is the radical interpreter as language user, apprised of various syntactic details, who can recognise the linguistic character of another person's behaviour and attitudes. These syntactic aspects of a language are not imposed on the surface of a speaker's practice by intentional re-description, rather they depend on cognitive structures that are causally implicated in acts of speech recognition and other linguistic capacities of speakers and interpreters. They are aspects of cognition that give form and character to a speaker's linguistic behaviour and knowledge.

Chapter Four

Now if the Davidsonian interpreter is supposed to recognise something as elaborate as the sentence structure of English, he must show how this can be manufactured on the basis of knowledge of certain a priori principles and observations of a speaker's behaviour. The empirical alternative is to base one's theory on linguistic data concerning speaker's intuitions, their language acquisition and development; as well as on constraints of psychological simplicity, psychological processing and on the data from language deficits. This suggests a rather different relation between linguistics and psychology; and it will be wholly illegitimate for the Davidsonian to claim knowledge of what is linguistically described and explained while repudiating this form of psychology. He must offer us his own route to a recognition of the grammatically well-formed sentences of a language, using no more than the ordinary intentional idiom. No-one doubts that we *know how to* pick out sentences in a language we have come across before, but interpretation theory is meant to give us an explicit account of *how* we do this. Once again it is obvious that one can know how to do this without knowing how one does it. What is rather less obvious is why the the account of how we do it belongs with the rational a priori theorising of the radical interpreter. What we need from Davidson is more of the details of how linguistic interpretation is carried out, and I shall turn to this in a moment. But notice that Davidson does admit that it is an aspect of interpreting a speaker that:

in passing from a description that does not interpret [his uttering of certain words] to interpreting description [his saying something by means of them], we must introduce a machinery of words and expressions (which may or may not be exemplified in actual utterances). (Davidson 1984, p126)

However, we need to know what that machinery is and what if anything it is working. Linguistic structures will have to be postulated in the course of introducing sub-sentential items, but it is the whole thrust of the interpreter's stance that if illumination is to be had we must not presuppose notions more mysterious than the explanandum. This is one

Chapter Four

of the attractions of radical interpretation: it promises elucidation of the notion of meaning without merely presupposing it. The same should hold for linguistic structures and the invocation of "machinery". Now when we try to elucidate the notion of linguistic structure and logical form we are offered very little. Either the radical interpreter presupposes structure (as in the syntactical case) or he refers us to the idea that it is the deductive structure of the theory that articulates the semantic structure of a language. But on the latter view we need to know whether the semantic machinery of the theory follows the psychological contours of the speaker's competence (or prescribes them); or whether it is purely an artefact of theory. As Martin Davies neatly points out, the idea of deductively revealed semantic structure:

...can be given a more a prioristic or a more empirical slant. For the structure to be articulated may be considered as an abstract (or mathematical) structure or as a psychological structure.
(Davies 1986, p131)

If it is treated in the former way I will still want to know how the interpreter has access to it. This is crucial since Davidson ends up by assuming that there is no asymmetry between speaker and interpreter. Whatever he says about semantic structure, I submit that he will be unable to account for syntactic structure, and that furthermore, his interpreter will have implicitly acknowledged empirically confirmable details about psychological structure when he identifies the sentences a speaker holds true; that these empirical details, concerning a speaker's cognitive equipment, are part the speaker's mind challenges Davidson's a priori conceptions of the mental. These conclusions have yet to be made out in detail but it is as well to have them in front of us as we examine further the workings of Davidson's semantic programme.

4.5) Putting the Empirical Constraint to Work

Now that we know what the interpretative requirement on linguistic theory must look like, we must see how it can be met. The claim is that it

is only by applying a theory to the undifferentiated evidence that the interpreter can extract an account of linguistic meaning from a speaker's linguistic behaviour. And now via the theory and the empirical adequacy condition we get an account of what it is for a sentence to take the meaning it does: for a sentence the speaker utters to have the meaning assigned to it by semantic theory the content-specification must pull its weight in the best interpretation we can give of that speaker. But now a potential problem looms for the notion of individual sentence meaning. If what it is for the sentence "Socrates flies" to mean, on the lips of the speaker, that Socrates flies, is for that assignment of meaning to pull its weight in an overall account that gives the best interpretation of his behaviour, then this formula will apply to every other sentence too. The sentence "Katy lied" will mean what *it* does on the same speaker's lips if *its* meaning assignment pulls its weight in the best interpretation of him. We will be reduced to giving the conditions for each sentence's meaning what it does in exactly the same way. There is no difference in the empirical conditions that support the assignment of meaning to the different sentences. The same condition makes those sentences mean different things. How can this be? Davidson's answer, if he can sustain it, will have to be that the theory of meaning, which as we know for him is a Tarskian theory of truth, will have to give different specifications of the meanings or interpretations of individual sentences. So let us now look at how this is to done.

The total evidence for the theory of truth consists in those intentional speech acts where the speaker reveals his behaviour with respect to whole sentences. What a Tarskian theory of truth has to do is to provide a recursive characterisation of truth for the speaker's language L which would enable us to specify truth conditions for his sentences. The condition of material adequacy Tarski placed on this project was Convention T, the condition that the theory should entail for each declarative sentence in the language a theorem of the form:

(T) S is true-in-L iff p

Chapter Four

where instances of this form (called T-sentences) could be obtained by replacing "S" with a sentence in L, while replacing "p" with a sentence in the language of the theory that translates "S". Obviously, Davidson cannot rely on translation to get at truth for L, so he relies on an intuitive (language-neutral) notion of truth to get at interpretation. What we might call Davidson's Convention D, would be as before, except to say that instances of (T) are obtained by replacing "S" with a canonical description of a sentence of the speaker's language L, while replacing "p" with a sentence of the interpreter's language with the same truth conditions as "S". To be interpretative, the truth conditions a theory assigns must be those which service the project of overall interpretation of the speaker. What an interpretational theory of truth then has to show is how the truth conditions of all and only the declarative sentences of L are determined as theorems. (Non-declaratives are assumed to be explained by some relation to declaratives described by a suitable theory of force.) The attributionist thought, here, will be that how the truth theory determines the truth conditional meanings of sentences in L provides a sufficient account of how the meanings of those sentences are determined; and since there is no other access to the facts described by a correct theory, the theoretical derivations provide the only sufficient explanation of how those meanings are determined. There can be no other account of how the meanings of sentences of L are determined for its speakers: the truth-theoretic explanation of how meanings are determined constitutes these facts. This idea will provide many of the conclusions we can draw about the Davidsonian programme, so let us suspend it temporarily and proceed to discover how the account of linguistic interpretation unfolds. Construction of the truth theory proceeds by finding a finite set of axioms from which all and only the T-theorems can be

Chapter Four

derived. Such an axiomatised theory for a speaker would satisfy Convention D and show how the truth conditions (meaning) of any arbitrary sentence of his language were determined. A further claim will be that this is a compositional account; that is, the theory will show the T-theorem for any given sentence to be determined on the basis of a semantic structure it describes for the sentence in question. Seeing a sentence as having semantic structure is a matter of seeing its semantic value, here, its truth or falsity, as depending on the semantic features of its parts and the inner structure of that sentence. The truth theory will display this by having the T-theorems for a sentence deducible from axioms that assign semantic properties to simple expressions occurring in that sentence together with clauses that deal with their modes of combination. The semantic features assigned to simple expressions must show their truth-affecting potential, and this is achieved when the axioms that assign them can be used in proofs of theorems that specify the truth conditions of whole sentences. But these aspects of sub-sentential semantics are not discovered by examining the empirical evidence: the fit between the theory and the physical facts it re-describes is at the level of the theorems. The compositional aspects of this account of meaning are theoretical assumptions that have to be justified somehow. As Davidson puts it:

A workable theory must, of course, treat sentences as concatenations of expressions of less than sentential length, it must introduce semantical notions like satisfaction and reference, and it must appeal to an ontology of sequences and the objects ordered by sequences. All this apparatus is properly viewed as a theoretical construction, beyond the reach of direct verification. It has done its work provided only it entails testable results in the form of T-sentences, and these make no mention of the machinery. (Davidson 1984, p133)

The question is do these theoretical notions have any 'indirect verification' ? To assess this we must confront the whole issue of the systematicity of semantic theory, and I turn now to this.

4.6) Formal Constraints on Truth Theory

A truth theory which really does describe how the meanings of a speaker's sentences are determined will get the facts about that language right. But what are those facts? It has been acknowledged that the only empirical facts to go on concern a speaker's use of whole sentences. And yet the truth theory tries to determine the truth conditional meanings of sentences on the basis of the semantic features of their parts. Why should it look into the internal structure of sentences if all that is required is to provide interpretations of sentences that will pull their weight in the reason-giving explanations? This is to question the need in an interpretational truth theory for a requirement of system.

First of all, what is this requirement of systematicity? It is a demand that we should see the meanings we attribute to a speaker's sayings as "determined by the contributions of distinguishable parts...of utterances, each of which may occur, making the same contribution, in a multiplicity of utterances." (McDowell 1978, p121) But now why should a theory aim to be systematic. Well, it might be argued, there are compositional facts about natural languages, and a systematic theory will reflect them. But this line of response is not open to an attributionist who sees linguistic facts about meaning as constituted by theoretically well-motivated descriptions of linguistic behaviour. At this stage all we can say is that the project of radical interpretation is systematic, not that the language is; and if anything, for the attributionist, the language will be systematic because radical interpretation is. But what we want to know is why radical interpretation assumes systematicity.

Is there is any less loaded way of expressing the need for system in interpretation? I think there might be, and that this may well be what Davidson is aiming at. The intuitive idea was that the interpretations we can give of someone's language seem to have some authority when they bring out the system in someone's language. This thought was then larded with the further claim that there was some system *in* the speaker's language which a correct truth theory would be capable of tracking. However, there is a different way to read the intuitive idea that suggests

that it is because interpreters are systematic in interpretation - seeking out many different points from which to fix on the meaning of some predicate expression - that the behaviour we are interpreting is given its distinctive linguisticity. The claim will be that the linguisticity of a subject's behaviour will consist in the interpreter's being able to deploy compositional resources to give a systematic interpretation of those mouthings.

This reply would, I think, serve the Davidsonian attributionist much better than the claim that the need for finite axiomatisation, or the need to explain speakers' capacities to understand any of potential infinity of sentences on a finite basis, requires systematicity. The first is redundant when we are dealing with a large but finite language, and the second should be no part of the attributionist's task given that he denies that there are any theory-independent psychological facts underlying linguistic behaviour. Instead, the systematicity constraint would be justified by the need to treat some parts of a speaker's total behavioural output as in essence linguistic. In this vein, recall Davidson's remarks about 'inducing semantic structure' on sentences on the basis of speaker's actions and dispositions to act. He can tell us in theoretical terms how we can do this:

...the kind of structure required seems either identical with or closely related to the kind given by a definition of truth along the lines first expounded by Tarski, for such a definition provides an effective method for determining what every sentence means (ie. gives the conditions under which it is true). (Davidson 1984, p8)

What Davidson cannot tell us is why such a theory is successful when it is at treating people's behaviour in such a systematic way. Why do we get away with it? The question will seem misconceived to the attributionist, but I think it is hard to resist the idea that there is something in virtue of which the theory is true when successful, and whose absence would explain why, in some cases, the theory would not work. The explanation would have to be given in terms of facts about the underlying psychological structure of language users. Of course such an answer is not open to Davidson, because his claim to exclusivity on the

mental, leads him to say that there could be no such psychological facts for successful theories to be about. We shall see.

Returning to the need for systematicity to support a claim that one is giving *linguistic* interpretations of behaviour, we can see that this feature of interpretations will provide a gloss on Davidson's reasons for thinking that the two component view of why a speaker holds a sentence true cannot be reduced to the single component of belief. No such Gricean reduction is on the cards so long as we must interpret so as to describe a linguistic system. The Gricean strategy would give us no insight into the character of a speaker's language. Furthermore, this claim for system in language may help with our earlier problem of how to separate out accounts of individual sentence meaning from the point of view of the interpreter, where we found that under the requirement of serviceability in the project of interpretation, they were all given exactly the same account in terms of their pulling their weight in making best sense possible of the speaker. Now we might be able to give systematic descriptions of their structure and their parts and, in terms of these, the relations they bear to one another. If these notions can be substantiated then so can the notion of sentence meaning.

Attributionists are also inclined to suppose that these facts about linguistic systematicity are what underpins our right to think of beliefs and other attitudes as having structured contents with conceptual constituents. The linguistic means of expressing and interpreting ourselves and others carves the contents of thoughts into concepts. This is the largely Fregean idea that the composition of a sentence serves to determine what thought it expresses and thus displays the conceptual elements involved in grasping that thought. For example, the order of composition matters when determining which of two thoughts is expressed by the ambiguous sentence, "Everybody loves someone". The different orders of composition reveal that on one reading the sentence contains the monadic predicate, "____ loves someone" and on the other, the different monadic predicate, "everybody loves ____". Neither semantically interpreted sentence contains the other. Thus in the first case, the predicate expresses the concept of loving someone which has the

property that everyone has it. In the second case, we have a predicate that expresses the concept of everyone's having love for something which has the property of being instantiated by at least one individual. Thus the two different thoughts expressed by each reading are explained by the different concepts exercised in each, and this is explained by our understanding of two different expressions wrought by different orders of composition of the same string of words.

However, this view may seem to raise a new threat. The systematic properties that radical interpretation ascribes to natural language prevented the reduction of linguistic meaning to speakers' beliefs and intentions. But in the light of the above view of the linguistic identification of concepts they might seem to threaten the other component of the two factor story; viz. the notion of belief. For the suggestion seems to be that the only understanding we can have of belief is that given in terms of the thinker's or the ascriber's potential for expressing a belief by a sentence in a public language. However, this view would not dispense with the belief-factor in the reasons for holding sentences, true, it would simply give a certain explanatory prominence to the linguistic side of interpreting a person's thought and action. Davidson, McDowell, and Dummett all subscribe to this thesis. Dummett has called the language-first doctrine the Fundamental Thesis of Analytic Philosophy (and more recently, the Fundamental Thesis of Linguistic Philosophy); it states that any account of thought should go via an account of the language in which that thought is expressed.

This thesis is quite compatible with the conviction that an account of language belongs to a larger account of making sense which presents a plausible description of the speaker's psychology. But if that is the case what would a denial of the Fundamental Thesis amount to? A recent example of its denial would be the thought-theoretic account of content given by Gareth Evans in The Varieties of Reference. In that work, Evans supposes that a proper treatment of the devices of linguistic reference rests upon a prior account of a speaker's capacity to entertain certain kinds of singular thoughts. The theory of thought will describe how particular kinds of singular thoughts relate the thinker to objective particulars. But

notice that although Evans may be reversing Dummett's explanatory priorities of thought and language, he still has need of both semantical and psychological theories. He will explain the sameness or difference in the senses of linguistic expressions in terms of the speaker's ways of thinking of the referents of those expressions. So he is not trying to replace the theory of truth that serves as a theory of linguistic senses by the theory of thought that informs it. The upshot is that on both of these non-reductive accounts of thought and language there is room for the two factor account of holding a sentence true. Moreover, the issue of whether we adopt thought-first or language-first theories is orthogonal to the issue of whether there is any significant semantic structure in natural language.

On my reading of him, Davidson wants to see system as an essential characteristic of language. The question will be whether he is entitled to do so. To assess this we must look at how the empirical requirement of interpretability and the formal requirement of system interact.

4.7) Truth Theory and the Physical Facts

When the interpreter begins to construct a recursive characterisation of truth for the language of a subject all he has to go on are the uninterpreted physical facts about that subject's behaviour. Prior to interpretation these facts cannot be described in content-specifying terms, and yet, even at this stage these facts must constrain the construction of truth theory. For if they failed to do so, it would be hard to see how a speaker's physical circumstances placed any constraints on how we could interpret him. We could simply dream up an analytic definition of his language and mind without reference to the particularities of his case. Clearly, then, there have to be some points of contact between truth theories and the kind of evidential facts which can support them. Let us now look at these.

To begin with there is the level McDowell calls "the hard physical facts" (McDowell 1978, p122). Facts at this level comprise "the structural properties of physical utterance-events" (Ibid) which enable us to construct a syntactic description of the language together with the complex

relations between a speaker's behaviour and his environment which will permit the initial and tentative identification of some of his behaviour as intentional action. Identifications of the latter sort will depend partly on making imaginative leaps to interpretations within the content-specifying mode of discourse; these intentional descriptions must be sustained in the developing interpretation, or eventually overturned. But identifications of the former kind are less easily revised: they are based on an assumed relation of match between the structural descriptions in the theory of syntax and "configurations observable in physical utterance-events" (Ibid.) It is not at all obvious what is supposed to be 'observable' in the physical events of utterance, but the tenor of McDowell's remarks suggest that syntactic properties of a language are just physical properties of sounds. Sounds, moreover, which are initially not even described as speech sounds. As McDowell puts it, we should think of a theory of language:

...as warranting systematic imposition of interpreting descriptions on a range of potential behaviour which would constitute speech in the language, thought of as describable, in advance of receiving the interpreting descriptions, only as emission of noise. (1977, p147)

What we need now is to know more about the interpreting descriptions supplied in the theory, along with some account of how they make contact with the hard physical facts. Earlier we noted that the only linguistic evidence we have is at the level of speaker's use of whole sentences, presumably because the hard facts about the noise emissions are only registered as linguistic behaviour once the interpreter has made a tentative identification of the partially interpreted attitude of a speaker's holding a sentence true. These facts about the use of whole sentences will also be facts about the constituent syntactic structure of the sentence, since this is also made available by the "hard physical facts". All we need to note now is that on the side of the theory, it is the theorems that record interpretations of linguistic behaviour, since it is at this level of Tarski-style truth theory that interpretation-serving assignments of meanings (truth conditions) are given to whole sentences. So it is at the level of its theorems that an interpretational truth theory makes contact with the

hard physical facts. Though it does this through the tentative identifications of a speaker's intentional relation to a sentence. This is a two stage process. Interpretation must effect a transition from the raw behavioural evidence to descriptions like:

(H): X holds "S" true when and only when p

and from there to: "S" is true-in-L iff p

which as we know is the target schema all of whose instances must be proved for the truth theory to satisfy Convention (D). But satisfaction of this condition will not yet amount to interpretation for we could have many truth theories for a language like English which allowed us to prove theorems of the form:

(T!) "Coal is a fossilised carbon" is true-in-E iff grass is green

Here Convention (D) is met just because we have a sentence on the right-hand side that is true under the same conditions as the sentence mentioned on the left, and so long as we are dealing with the material biconditional this will be a true consequence of candidate truth theories in which it can be proved. But it is obvious what we should say here. The evidence for an interpretational truth theory does not concern any one T-sentence considered in isolation. We have to look at the whole range of T-sentences provable from a theory with the resources to prove this one. So long as any such theory is systematic, it will pair the expressions it describes in the left-hand sentence with the same expression each time on the right-hand side. Indeed this correlation will be recorded explicitly in the axioms which serve as premisses for derivations of T-theorems for sentences containing those expressions. Now in such a case it is unlikely that all of the T-sentence pairings involving the expressions and assignments used in (T!) will be true. To try to bring the theory back to conformity with the evidence we would have to violate the requirement of systematicity by allowing the very same expression to be given different

interpretations on different occasions of use. Not only would this spoil our chances of interpretation, by ruling out the possibilities of our finding multiple points from which to get a fix on any particular expression; it would also violate the holistic interdependencies whereby the correct interpretation of each speech act is given in the light of all others. It is the totality of T-sentences derivable within any T-theory that we must use to test the interpretation a theorem gives of any particular sentence. The legitimacy of a given truth theory cannot be tested by one or another T-sentence: we must test the whole theory and all its logical consequences. Holism and systematicity are mutually reinforcing in conspiring to constrain the choice of genuinely interpretative theories of truth.

4.8) Systematicity and Epistemological Holism

Interpretation provides our only understanding of the mind: minds and their contents are what we attribute to people in the course of interpreting them as agents. So the formal and empirical constraints on interpretation provide the contents of the mind with whatever significance they have. The crucial importance of these constraints consists in the sorts of intentional notions they make possible. Linguistic interpretation plays a central role in the strategy of the interpreter so we need to know exactly what is made available by linguistic interpretation of someone's language. I have suggested that on Davidson's account the factors which make it possible are systematicity and the holism of attitude ascription. We need now to see just how they do this and precisely what range of linguistic meanings they prescribe.

Systematicity requires us to discern a common feature in the interpretation of sentences which contain the same linguistic expression. Holism requires us to adjust the interpretation of each sentence uttered to the interpretation of all others. The holistic constraint is entailed by the fact that the evidence for sentence interpretation will be the speaker's attitudes and actions directed towards sentences and utterances; in this way sentence interpretation must be sensitive to the possibilities of belief ascription. Systematicity is needed so that our attempts to interpret, in

taking in all other aspects of a speaker's mental life, should be more than just a case of propositional attitude ascription; they must be interpretations of a linguistic system.

It is important to notice, here, that postulation of system does not entail holism: the holism Davidson finds in the facts about meaning are imbibed through the evidential-cum-constitutive conditions for fixing meanings and beliefs. What system does provide us with is a means of projecting meanings from the evidence of sentences actually used by the speaker to those in the range he could potentially utter. However, holism disrupts the stability of this picture. Any projected interpretation of new cases that are tested against speakers' responses may well be disconfirmed if they do not sit well with the best interpretation of the speaker's total behaviour when we incorporate these additions to his repertoire. These cases can lead to revisions in the systematic patterns already offered. Thus projection is a risky business - there may be many ways to project all equally compatible with the total evidence available at a given time. But not all of them will stand up in the light of further behavioural evidence (the hard physical facts and the initially conjectured attitude ascriptions). Here we see why it is holism that breathes life into the system. Systematic descriptions must be at the same time rational explanations.

But of course the holistic possibilities must be developed in league with some type of system, so where there are choices available to the interpreter we want to know both what range of systematic descriptions he can work with, and on what basis he chooses between candidates within that range.

Supposing for a moment that there are a number of possibilities, we can see how more and more confirmations of projected interpretation will deepen his confidence that he has the right view of the actual fragment of the language he has interpreted. And it is this increased confidence in giving interpretationally sound linguistic descriptions of a speaker's utterance that will lead the interpreter to attribute a language to a speaker characterised by the semantic features and structures of the system he is using. But of course he can only begin to project to new cases by reading structure into the sentences he encounters in speaker's practice.

What he has to search for is a system of semantic features and internal structuring on which to base an inference from old cases to new: from the facts about combination as he sees them to the possibilities of recombination permitted by the speaker. Increasing one's confidence that one had latched on to, what we might call, a good system for interpreting someone's language will enable one to individuate that person's beliefs more precisely; in particular it will enable one to detect false beliefs of the speaker.

So how does the interpreter construct a truth theory for someone else's language? It works like this. The common features discerned in interpretations of sentences sharing an expression will be semantic features from which we can assemble truth conditions for whole sentences. These semantic features will be assigned to expressions by the axioms of the theory, and they will serve, in concert with similar axioms, as premisses in the deduction of theorems that specify the truth conditions of sentences containing those expressions. So a systematic truth theory must be structurally insightful in the sense that it must assign to sentences truth conditions on the basis of their semantic parts and their internal structures. However, different axiomatisations will lead to different structural insights and so what we need to know are the parameters within which we can give axiomatisations of a truth theory for a speaker's language. The remarks so far have given only the briefest indication of what these semantic features and structures should be like. But when it comes to the question of what Davidson thinks of as fixing semantic structure, we find that he has very fixed ideas. Davidson believes that whatever the choices of axiomatisation are, they must all take the form of a Tarskian semantics. Let us quickly sketch out this assumption.

The sentences held true by a speaker provide crucial evidence of theorems which belong to the set of all theorems we want to prove. The theorist's job is to select the smallest set of axioms from which to derive those T-sentences. To do this the theory must be structure-discerning. For it is only by thinking of sentences in terms of parts and wholes that one can see the projected truth conditions of new sentences as being based on the truth conditions of the old ones. With the right formal conditions the

truth conditions for new sentences can be computed as functions of the permissible permutations of the semantic parts of the old ones.

This was the justification of compositional structure given by Tarski in the original semantic definition of truth. Tarski wanted to give a uniform semantical account of the logical operators, but the semantic treatment of them as truth functions from the set of truth values onto itself broke down in the case of the quantifiers. The truth values of logically complex sentences with a quantifier dominant could not be treated as a function of the truth-values of their parts, for their parts were not sentences and thus were not truth-evaluable. Tarski solved the problem of uniformity while retaining the idea of the operators as functions by introducing the notion of satisfaction conditions to give the semantic values of not just sentences but also what he called open sentences: sentences with the quantifier prefix removed. An open sentence was one which could be satisfied by an infinite sequence of objects by treating the unbound variable in the sentence as a temporary name for one of the objects in that sequence. A closed sentence with all its variables bound is one that is satisfied by all sequences or none at all. Tarski defines truth as the first of these conditions. In this way he is able to treat the satisfaction of logically complex sentences uniformly as a function of the satisfaction conditions of their parts, while still being able to account for the truth of whole sentences. The importance of this development for our purposes was that it led Tarski to look for structure inside the atomic sentences, normally treated as wholes on accounts of the propositional constants, and this gave rise to the possibility of a structure-discerning theory of truth.

Turning to an axiomatic theory of truth, we have to import another notion of subsentential semantics, namely the reference of singular terms. Tarski neglected these since his concern was with the variable in quantified sentences. But by adding this notion to those of truth, satisfaction and combination we have the semantical materials to construct a finitely axiomatised theory of truth. The axioms will look like this:

(R1) "Socrates" denotes Socrates

(S1) An object satisfies "flies" iff it flies.

(C1) A sentence combining a name *n* with a predicate *F* is true iff the name denoted by *n* satisfies *F*.

From axioms like these together with a canonical set of proof procedures we can prove theorems like:

(T1) "Socrates flies" is true-in-L iff Socrates flies.

All such T-sentences must be deductively derived from a finite set of axioms that deal with expressions and constructions that occur in sentences on their left-hand sides. The axioms selectively called upon to prove any particular T-sentence for a given sentence of L will reveal the semantically significant constituents and the logical form the theory assigns to that sentence. In this way, the axiom structure of the theory shows how the reference (interpretation) of an expression contributes to the truth conditions (interpretation) of sentences in which it occurs, thus conferring on it a significant semantic role. And it can do this because the canonical proofs of T-sentences reveal the semantic structure the truth theory assigns to object language sentences.

But do these semantic features and structures characterise properties of a speaker's language? To answer this we must turn from the formal constraints to the empirical ones again, and ask what makes such a truth theory a theory of interpretation?

An acceptable truth theory will be one whose T-sentences are confirmed by the interpreted evidence about the sentences a speaker holds true (adjusted for false beliefs). The object of linguistic interpretation is to provide the right assignment of truth conditions and these can be thought of as the acceptable T-sentences at which a correct theory must aim. So the important part of the semantic enterprise, so far as interpretation is concerned, is the empirically testable output of truth theory. Indeed this is the only level at which the truth theory makes contact with the evidence since the only linguistic evidence there can be is evidence about the use of whole sentences. However, that level of evidence is meant to be revealing

about the syntactic structure of the sentence, at least if we can believe the story that says syntactic structuring is just a property of the physical facts of utterance. Now we must imagine the interpreter constructing a truth theory that employs the notions of reference and satisfaction to derive all and only the acceptable T-sentences for a given language. But what is it to say that a given T-sentence has adequately characterised a speaker's language? All empirical adequacy requires is that it should get the truth conditions right: that the pronouncements of the theory should meet the Propositional Attitude Constraint. But of course the formal constraints put conditions on notions below the level of the sentence. The truth theorist has to supply axioms that specify the references of simple expressions in *L*, together with axioms or rules that specify the semantic impact of combining them in particular constructions. Is there any empirical evidence to support the axioms? Not directly, as we know, but they might be said to be indirectly evidentially supported just in case they enable us to derive the right set of T-sentences. The T-theory would be anchored to the facts not at the level of axioms but at the level of theorems. The deductive means for getting from axioms to theorems would appear to be irrelevant so long as they permitted the derivation of empirically adequate theorems. (In fact very little is needed here: minimal logic would suffice.) But it is the internal workings of the truth theory that allow us to read semantically significant structure into sentences: the semantic structure of a sentence is not exhibited in its T-sentence but in the structure of the proof by which it was derived. All the evidence the radical interpreter has to go on is a pattern of truth-values assigned by speakers to sentences of the language under specific conditions given by the beliefs we can ascribe to them about what is the case when they assent to those sentences. It looks as though the theoretical notions needed to describe how those truth-values are determined can only be confirmed from the theorems up. But now, if there is any other way of tracing back up to a different set of axioms and proofs confirmed by the very same theorems then we have a distinction without an empirical difference.

Does this last claim show that notions of structure and the references of sub-sentential expressions have no empirical content? Well, we have

Chapter Four

to be careful here. What we know is that the empirical and formal adequacy conditions do not coincide. The formal conditions govern the deductive direction from theorems to axioms dealing with the inner composition of sentences, whereas the empirical conditions relate to the output of the theory at the level of theorems. But this is not to claim that there is no connection between the two types of conditions. Perhaps, the empirical content of an adequate truth theory flows back up through the derivational structure of the theory, transmitting content to the axioms indirectly through the multiplicity of connections they enjoy with the testable T-sentences they give rise to. This would allow us to see each T-sentence containing a given expression as providing a point from which to fix the content of the axioms that deals with it. Just so long as we are systematic there will be good evidence to think of a word as having the semantic feature we use to arrive at more and more interpreting descriptions of sentences containing it. It might be fair to say that the semantic properties of words that the axioms record are abstracted from the semantic properties (truth conditions) of sentences in which they appear.

Now we can see a way of putting the constraints together: T-theories should generate all and only the correct T-sentences for the language in question; evidence at the level of these output theorems would confer empirical adequacy on the formally correct theories that give rise to them. However, a major difficulty arises at this point. What are we to say of empirically adequate theories that implement the formal constraints in different ways. Whenever there are different sets of axioms and proof-theories that enable us to derive the same set of T-sentences as theorems they will count as equally adequate but empirically equivalent truth theories. Does this make it a theoretical insight rather than an empirical discovery that the truth values of a speaker's sentences are determined by the semantic properties of their constituents under some arrangement? Then the claim to be able to find semantical features of subsentential expressions depends less on the linguistic practice of the speaker and more on the theoretical principles that impose structure on sentences as a formal requirement of theory. Notice that this answer will give us no clue

as to how the empirical consequences can put pressure on the form of the theory by transmitting falsity from theorems to axioms in incorrect theories. There is a world of difference between finding that there are many axiom pools from which to prove the target biconditionals, and finding that a disconfirmed T-sentence must lead to a shift in the resources from which it was derived. In this way, the upwards transmission of falsity from theorems to axioms forces revisions in the theory and thus gives rise to a shift in intra-sentential insights on the language. But the trouble is that empirical pressure of this sort is merely pressure to change, but it comes without instruction of what to change. (Borrowing a phrase from work on the diagnoses of error in computer programs this could be called "The assignment of blame problem".) We could revise the proof theory for example, which could lead to a change in the proof structure of theorems; or alternatively we could revise some of the axioms. Davidson denies us the chance of changing the logic since he says we must read into the language we interpret as much of our quantificational structure as is necessary to carry out the interpreter's task. Besides, the choice of logic is not important within the theory since even minimal logic will enable us to derive the biconditionals. The revisions must be made to the axioms. These include axioms dealing with specific constructions so revisions to the structurally-insightful descriptions we give to a language are still possible. Ideally, what we want to know is how much choice there is here in how we make the revisions. Care must be taken since we test T-sentences not singly but as a whole, so we must adjust the axiomatic resources of the theory in such a way that the revised theory is assured of empirical adequacy. But even still, choices may remain, and the point is that there will be no empirical way to choose between them.

Does this mean that there is no empirical content to the structure-discerning aspects of truth theory? There is as much content as the evidence for T-theorems confers upon them. I suppose the follow up to that answer would be a question about the likelihood on this framework of massive indeterminacy within the structure of individual sentences. How likely is it, when empirical push comes to shove, that there will be a

huge variety of axiomatisations that will all guarantee the right set of theorems as output? Davidson thinks the fact that we have to take into account the totality of T-sentences and that we have to bring to bear on that totality a systematic truth theory in the Tarski-style, means that a good deal of indeterminacy will be brought under control. (see Davidson 1984, p228) But this answer highlights the fact that it is the theorist's assumption of system and not the empirical facts that give rise to the logical form of a sentence. It is the deductive apparatus of truth theory that induces structure on the sentences a speaker uses. Davidson speaks of the deductive apparatus as theoretical machinery that makes no showing in the verifiable parts of a theory. He says that a T-theory has done its work provided that it entails testable results in the form of T-sentences. So should we conclude that the intra-sentential insights are just empirically irrelevant? Would any means of getting to the right theorems suffice? It seems to me that Davidson should be in a quandry here. It is the assumption of system that allows him to persevere with interpretation when faced with such bare evidential materials, and the assumption seems like a purely a priori speculative principle. But once again one wants to ask why does the assumption work when it does? Why do the revisions of T-theories in the light of disconfirmed projections sometimes work out and sometimes not? Qua attributionist, Davidson has no answers to these questions: we know a priori when the theory is successful because we know a priori when the empirical conditions support such theoretical descriptions.

The question was how much does the empirical evidence constrain the choice of axiomatic truth theories, and why? On my reading of him, Davidson, tells us that the only pressure exerted on the choice of axiomatisation comes from the formal requirements of system and the interpretative demands of the Propositional Attitude Constraint. He proclaims:

A theory of truth thus reconciles the demand for a theory that articulates grammatical structure with the demand for a theory that can be tested only by what it says about sentences. (Davidson 1984, p133-4)

But how much empirical life do these two requirements, acting in concert, breath into the semantic structures of sentences and the axioms of truth theory from which we derive truth conditions for those sentences? The Propositional Attitude Constraint is itself a source of indeterminacy: we know that the anomalous mental leaves room for more than one configuration of internal states and intentional actions that enjoy a fit with the patterns in a normal person's behaviour. So this will lead to indeterminacy about choice of truth theory. Secondly, we know that the references of expressions recorded in the axioms enjoy no direct fit with the empirical evidence of linguistic practice, and so where different reference assignments to words are compatible with the same systematic assignments of truth-values to whole sentences, there will be indeterminacy of reference. And lastly, we have Davidson's merest hint that in choosing the Tarski-style format to realise his assumption of systematicity in linguistic theory, he will have cut down on a great deal of indeterminacy to do with logical form. Indeterminacy will arise whenever there is a range of formal options within a theoretical setting that make no empirical difference whatsoever. The crucial question is to determine just what tolerances we have to work to when formulating the means to implement an interpretive truth theory. The interpretive needs are met by "showing how the theory can be supported by relating T-sentences, and nothing else, to the evidence" (Davidson 1984, p223). Meanwhile, at some places in his writing, Davidson seems to suggest that anything not mentioned in the T-sentences themselves must be considered purely theoretical and entirely lacking in empirical content. Speaking of reference and satisfaction Davidson says:

these notions we must treat as theoretical constructs whose function is exhausted in stating the truth conditions for sentences. Similarly, for that matter, for the logical form attributed to sentences, and the whole machinery of term, predicates, connectives, quantifiers. None of this is open to direct confrontation with the evidence. (Davidson 1984, p223)

Chapter Four

But in other places, Davidson expresses confidence that there is something in common between empirically adequate truth theories that purport to characterise a given language.:

reasonable empirical constraints on the interpretation of T-sentences....plus the formal constraints will leave enough invariant as between theories to allow us to say that a theory of truth captures the essential role of each sentence. (p224)

and

what is invariant as between different acceptable theories of truth is meaning. (p225)

So some aspect of the meaning or the role of any sentence is not itself crudely empirical or behaviour at all, but a property of the theory. Now given this conception of meaning it becomes of immense importance in linguistic terms to know where the indeterminacies fall and where they do not; and where they do not, why they do not. Answers to these questions mark out the territory left to Davidson for the notion of meaning or interpretation, and the importance of these issues cannot be overstressed:

Indeterminacy of meaning or translation does not represent a failure to capture significant distinctions; it marks the fact that certain distinctions are not significant. (Davidson 1984, p154)

This is an important admission on Davidson's part and it sets the stage for his resulting position which I shall call **Semantical Minimalism**. Minimalism is what Davidson's theory of linguistic meaning leaves room for, and we can only appreciate the nature of this doctrine by seeing what survives on Davidson's conception of linguistic interpretation.

Chapter Five

Indeterminacy and Meaning

5.1) Meaning and Logical Form

I want to start examining Davidson's minimalist conception of meaning by considering the indeterminacy of form. This is a crucial issue for Davidson given his use of the assumption of systematicity to identify linguistically interpretable material. For this reason, I want to raise the question of form in the light of a powerful challenge to structure-discerning theories made recently by Crispin Wright (1981, 1983, p74, 1986a pp31-44, 1987 pp204-39). The question Wright raises is whether there is any good reason for insisting that a Davidsonian truth theory should be structure-discerning in the way described above. If Davidson's answer is given in terms of the uses to which such theories can be put, Wright will want to know more precisely what those uses are and whether they can be met by non-structure discerning theories.

Davidson wants theories of truth to serve as theories of meaning by providing correct interpretations for each of the infinitely many declarative sentences of the language in question. They can do this by providing the deductive means for generating infinitely many theorems that make acceptable assignments of content to each of those declarative sentences. The acceptable content-assignments will be those that pull their weight in making best overall sense of the life and conduct of speakers of that language. Now the question Wright asks is why such a theory should postulate a semantically articulated structure for the language, given that it fits the empirical evidence from the theorems upwards and is empirically testable only at the sentential level. Why should a theory that is only required to deliver the right T-sentences be thought to tell us anything about the structure of the object language sentences themselves? Davidson's reason for thinking that it must seems to rest on an assumption that any theory capable of yielding a totality of T-sentences fitting the evidence, and extendable to infinitely many other cases, must invoke a semantic machinery for deriving those T-sentences from a finite set of axioms. The axioms will deal with each of the finitely many expressions and constructions in the language, and will entail for each of

Chapter Five

the infinitely many declarative sentences, a meaning-specifying theorem whose proof structure shows how the meaning of that sentence depends on its composition. In this way, the canonical derivation of each T-sentence will record a structure assigned to the object language sentence it contains, and hence, so long as a truth theory is finitely axiomatised it will yield insights into the structure of the language to which it applies. But to accept this claim is not to accept that it is necessary to construct a finitely axiomatised truth theory. So why does Davidson think that it is necessary? Perhaps, it is because he supposes that there is no other way to state finitely a theory for an infinite language: only a finitely axiomatised theory is capable of yielding all the right deductive consequences for an infinity of cases; and any such theory will provide insight into the structure of the language. It is this claim that Wright casts doubt upon in the homophonic case - the case in which the language for stating the theory is an extension of the object language, minimally enriched by the addition of certain key semantical terms. Here, the T-theorems are merely disquotational, and Wright suggests that "provided we have a recursive specification of the syntax of the (declarative part of the) language", we can arrive at the meaning-specifying theorems for every declarative sentence of the object language by means of a theory which:

merely stipulates as an axiom every instance of the schema:
A is T if and only if P,
where 'P' may be replaced by any declarative sentence of the
object language and 'A' by the quotational name of that
sentence. (Wright 1987, p221)

This theory is not finitely axiomatised, but it is finitely stated. It accomplishes the same task as the finitely axiomatised theory, since in conjunction with the recursive syntax, it provides an effective means for arriving at the empirically adequate, and hence meaning-specifying, T-theorems for every declarative sentence in the language. But it does this without revealing structure in the sentences on the left-hand sides of the biconditionals; and if we persist in thinking that the canonical proof of a T-sentence reflects the logical form the theory assigns to the sentence,

Chapter Five

then these sentences are assigned no semantic structure at all. (It will however have to find a syntactic articulation for these sentences which as I suggested above will require detailed empirical inquiry of something other than just speakers' linguistic behaviour.) The theory would still present an interpreter with the deductive means to arrive at information sufficient for understanding every (declarative) sentence of that language. So it would seem that Wright has offered Davidson a theory which meets the needs of interpretation theory without appealing to "the machinery" that articulates semantic structure. The moral according to Wright is clear:

.....the ambition to describe information which would suffice for mastery of a particular language may impose certain constraints on the form taken by the theorems of a theory of meaning, but it imposes no interesting constraints on the mechanics of the theory. (Ibid. p212)

This is a powerful point when combined with Davidson's thought that language is by and large a theoretical construct attributed to speakers by interpreting their behaviour. For if language reflects the properties the theory assigns to it, and it is not a necessary condition of an empirically correct truth theory that it should attribute logical form to speakers' sentences, then a natural language is subject to the most extensive indeterminacy of logical form. Of course, Wright is not proposing any such indeterminacy, he is merely pointing to a deficiency in Davidson's programme which leaves it unclear why one should seek to give a structure-discerning theory of meaning in the first place.

Should the Davidsonian accept the moral Wright draws from this example? The conclusion seems rather overstated, for it concerns the homophonic case alone, and if we are to take Davidson seriously when he insists on the global nature of radical interpretation, then he can never afford the unthinking complacency of a homophonic translator. It is only a modest theorist, like McDowell, who assumes that the concepts and contents used on the right-hand sides of T-sentences are exemplified in the sentences mentioned on the left. Here we can always use the sentence itself to state its own truth conditions, and against this modest proposal

Wright's point is quite telling. But why should it apply to the project of the radical interpreter too? Surely Davidson will insist that we have to earn the right to assume that what someone's sentences mean on his lips are what they mean on ours. All that the homophonic case comes to, for Davidson, is just a familiarity with the *sentences* the speaker uses, but not with the beliefs speakers have or the meanings they attach to those sentences. Constructing a homophonic truth theory for a speaker of my linguistic community is the first move in constructing a theory of his beliefs; but the initial choice of theory is a conjecture that may have to be revised in the light of an evolving interpretation of that person's behaviour. The use of a homophonic truth theory, then, is never a guaranteed prior assumption of interpretation: it is a hypothesis that may be vindicated by a successful interpretation. So perhaps it is the process of radical interpretation that calls for a compositional theory of meaning. I will examine this claim in a moment. But leaving the radical interpreter aside, it is worth noting briefly the wider implications of Wright's worries.

The real concern is with the underlying motives and ambitions of those who insist that a Davidsonian theory of meaning should describe semantic structure within sentences. The homophonic case serves to remind us that the demand that semantic theory should assign structure is additional to the demand that the theory should be finitely statable while being able to account for the meaning of the infinitely many sentences belonging to a given language. So what extra requirement is a structure-discerning theory being asked to meet in the homophonic case? Or as Wright puts it, what project would call for the provision of a compositional theory of meaning (by which he means a finitely axiomatised theory in which the meanings assigned by theorems to sentences are shown, in the derivations of these theorems, to depend structurally on the meanings assigned by the axioms to their parts)? If the ambition is merely to provide a complete description of the meanings a speaker could attach to any of the potential infinity of sentences he could use and understand, there will be no need to come up with a compositional theory. So what additional requirement is it that calls for

the postulation of semantic structure and the provision of a compositional account of language? **Wright's Challenge**, as I shall call it, is to come up with a project whose successful completion calls for a compositional theory of meaning. The homophonic case shows quite clearly that we are looking for a theory of meaning to do something more than just describe the meanings of a speakers' sentences. Wright is interested in whatever else it is that proponents of Davidsonian theories of meaning see these accounts as providing. For whatever it is, it must be the underlying reason for thinking, and so the basis for defending the idea, that a Davidsonian truth theory provides the right form for a theory of natural language meaning.

Two conceptions suggest themselves, both of which Wright regards as highly suspect. One involves seeing the theory of meaning as an attempt to describe the inner workings of the language, quite independently of its speakers' knowledge of it. The other sees an important role for the theory of meaning in explaining the linguistic capacities of actual language users. The former view treats compositional structure as a mathematical property of the way the autonomous language works. The latter view treats compositionality as reflecting the psychological form of speakers' competence. Despite their differences Wright is opposed to both views, for the Wittgensteinian reason that they each in their different ways try to import what Wittgenstein would regard as a spurious objectivity into the notion of a language. Both types of theorist believe that the correct axiomatisation of a theory of meaning for a particular language reveals something of the essential character of the language in question, and that the axioms codify "information which systematically settles the content of so far unconstructed and unconsidered sentences". (Wright 1981, p112) So both types of theorist would, in their different respects, consider the project of constructing systematic theories of meaning for given natural languages to be the right way to capture certain crucial facts about them; the determinate character of a language is what one aims to describe in the semantic theory. In one case, those facts would be autonomous; being both speaker-independent and beyond the epistemic reach of the non-specialist. In the other case, the relevant linguistic facts would be determined by

(though they would not be identical with) certain psychological facts about language users, facts which although quite firmly *dependent* on the speakers' epistemic reach, would not be explicitly known to ordinary speakers but only to the specialist.

While I share Wright's reservations that any Platonistic conception of the former sort will create an epistemologically intolerable divide between meaning and understanding, because it treats the 'facts' about the language independently of what speakers of that language know, I see less reason to doubt the intelligibility of the second position. What this position calls for is a notion of a speaker's implicit or tacit knowledge of the details of the formalised theory of meaning. Without this there is no way of saying how it is that a structure-reflecting theory of meaning, most of whose details are not explicitly known to the ordinary language user, can nevertheless play a central role in explaining his linguistic capacities *and* what he ordinarily knows about his language. On this conception, the theory of meaning will play a part in the psychological explanation of a speaker's cognitive capacities, and not just the rational explanation of a speaker's linguistic behaviour. For this reason it departs from the true Davidsonian picture.

There are many things to be said about the psychological explanation of linguistic competence, and many objections to it that Wright raises, but these are topics for the second half of the thesis. So what I want to consider now is whether Davidson's radical interpreter has to call upon a structured semantic theory to pass from descriptions of a speaker's behaviour that do not interpret to descriptions that do?

5.2) Why Should Davidson's Semantic Theory be Structure-Discerning?

Davidson's postulation of structure in language must be reconciled with his acceptance of the following two claims. When we begin to interpret speakers of L we bring to the task a theory of linguistic interpretation that provides insights into the structure, words, word-

meanings and references of items of L; these are what make for the distinctively linguistic materials of the interpreter. However, these theoretical insights into the internal workings of a language are merely brought to bear on the evidence by the interpreter but can be given "no independent confirmation or empirical basis" in the data for linguistic interpretation. How can we hold on to the claim that semantic theory must be structure discerning, as opposed to the merely *de facto* claim that we tend to carry out interpretation in a systematic way. If systematicity is a requirement of an adequate truth theory, it needs to be justified.

Structure is a delivery of the theoretical description of a language, but why should this be a desideratum? One answer that suggests itself has to do with the project of rational reconstruction, or what is really for the attributionist, rational postulation. Only the theory of interpretation, and in particular that part of it which deals with linguistic interpretation, brings to light the facts about meaning. A physical sound or mark, like a mental state, is meaningful only as described; so the theoretical description we give it is very important. The principles of the theory that does the describing will constitute all the linguistic facts. But why should such a project of rational reconstruction invoke principles of semantic structuring and not just the infinitary axiom schema? I shall now try to say why.

Well, what would it be for our semantic theory to be systematic? How should it work? First, we would presume that there could be configurations in the language L of sentences, S_1, \dots, S_n , and S for which someone could come to have knowledge of the meaning of S as a result of knowing the meaning of each of S_1, \dots, S_n . Then we could, by constructing a finitely axiomatised truth-theory, show that the only theoretical resources needed to derive the T-sentences for each of S_1, \dots, S_n would suffice for the deductive derivation of the T-sentence for S. Apart from the standard logical resources of a truth-theory, the derivations would depend, in particular, on the axioms dealing with the semantic structure and the semantic properties of parts of S_1, \dots, S_n . The assumption would have to be that these derivations included the axioms called upon in the derivation for S, and so S had parts in common with

all of S_1, \dots, S_n and shared a common structure with at least one (or one sub-part of one) of them. But these configurations of S_1, \dots, S_n, S give rise to two very different versions of this structure and parts account. And depending on the nature of the cases, the part-whole structures could be implemented by theories which reveal different degrees of structuring. Here are the two cases.

For any of these configurations, it will be possible to suggest that parts of S are contained somewhere in the set S_1, \dots, S_n . For some configurations of this sort, this could be true in virtue of S 's being merely a logical compound of all of S_1, \dots, S_n , giving recursive application to a logical operator already in use in at least one member of the set; eg. S could be ' $A \& (B \& C)$ ' and S_1, \dots, S_n , could be $\{A \& B, C\}$. This would not yet oblige us to discern, or describe, sub-sentential structure, though it would require the logical structuring of sentences into molecular and atomic statements. But for at least some configurations of S_1, \dots, S_n, S , it is plausible to suppose that S is an atomic statement, composed, in the main, of semantically primitive expressions. An expression is a **semantic primitive** "provided the rules which give the meaning for sentences in which it does not occur do not suffice to determine the meaning of the sentences in which it does occur" (Davidson, 1984, p9). So if, for the case of atomic S , we can understand it in virtue of understanding each of S_1, \dots, S_n , then the truth-theoretical description entails that S and the members of S_1, \dots, S_n have semantic primitives in common. These will be expressions each of which is given a meaning, or semantic role, by an axiom in the set from which we can derive all the T-sentences for S_1, \dots, S_n and S . Once the theorist finds a set of axioms which meet these requirements, the T-theory to which they belong will have established a stock of semantically relevant sentence-constituents whose structured permutations in accordance with the principles of construction will take us beyond the original set of meaningful sentences to a host of new ones. In this way the detection of sub-structure in the original sentences treated permit us to extend this evidential base. The thought here being that, when seen through the eyes of the theory, the constituents and structures

of novel sentences are already present in the sentences of the evidential base just sampled.

Now, if we take our opening presumption that there are such configurations to mean that: "it would be possible for someone to proceed by rational inductive means from knowing what S_1, \dots, S_n mean to knowing what S means" (Davies 1981, p56), then our finitely axiomatised theory would provide a rational explanation of the means of achieving this. (However, this would not be the means used by any ordinary speaker.)

Notice that such a theory would also explain why, when we revise the belief we attribute to a speaker, and hence the truth-condition we assign to the relevant sentence S , when the speaker holds that sentence true, we have to make corresponding adjustments in a number of beliefs and truth conditions ascribed to the speaker with respect to other sentences he holds true, to secure conformity with the behavioural evidence, on the one hand, and the rationality, coherence and consistency of his beliefs, on the other. In the sort of configurations just mentioned, revision to the truth-conditions of S would require a change in the T -sentence, and hence a change in its canonical proof. If this means changing the axioms then this will force alterations in the axiom pool for S_1, \dots, S_n and hence revised derivations for some, if not all of them, thus changing their meanings.

There may be some evidence for such revisions amongst the corpus of sentences held true. And the finitely axiomatised truth-theory would provide the best interpretation of any such data; ie. the best rational reconstruction of what is going on here.

Now we can answer the original question: why should we assume, as we did to begin with, that the semantic theory should be systematic? The answer is: because it provides a rational reconstruction of an ideal speaker's knowledge of meaning; the means by which he could arrive at the meaning of any of the indicative sentences of his language.

Davies explains the reconstruction as follows. The semantic theorist may judge there to be sufficient systematic similarity of meaning assignments (truth-conditions) amongst sentences which are syntactically similar to

....enable a rational agent (relevantly like himself) to reason inductively from knowing what certain sentences mean to knowing what other syntactically similar sentences mean. It is this judgement about the possibility of self-conscious, reflective projection of meanings which encourages the attempt to provide a theory of meaning which not only delivers the correct meaning specifications but also reveals how the meanings of sentences depend on upon the occurrence of particular syntactic constituents (roughly: words and ways of putting expressions together). (Davies 1981, p56)

Davies takes his Structural Constraint as a constraint on theories of meaning, whose descriptive job it is to reveal the structure of the language in a Platonistic sense of this. A finitely axiomatised semantic theory will implement this task, but it leaves it as a further question how the understander actually tackles the task.

I shall depart from Davies's description here by suggesting that the full-blown attributionist is making a decision, and not a judgement, about structure being, not revealed, but imposed on the language; and I shall equate the rational agent with the interpretation theorist himself. Of course the decision to impute semantically relevant structure to the data may not work out if the choice of structures on which the projections are based is not borne out by further evidence from the speaker's corpus. But if the imposition of structure can be sustained, the account will constitute a rational explanation of the speaker's practices in terms of rationally derived knowledge of the truth conditions of all the indicative sentences of his language. This is what an explicit theory of what the interpreter achieves is supposed to provide. Furthermore, such a theorist can proceed by rational means to an understanding of novel sentences without any further empirical investigations: this is a purely philosophical project. Complete knowledge of the natural language a speaker uses (or as complete a knowledge of it as the vocabulary and syntactic competence of our informant allows for) can be arrived at by a priori inquiry.

This suggestion corresponds closely to the strong interpretation of Davies' Structural Constraint on theories proposed by Wright in his 1987,

p215, but with stronger attributionist emphasis here on the *imposition*, instead of the *revelation*, of linguistic structure.

We have now arrived at what I believe to be the correct resting place on this matter for the attributionist. He must settle for a theory-internal justification of structure. But since the facts about the theory constitute the facts of the matter about meanings and minds - linguistic like mental events are only what they are as described - it is legitimate for the attributionist to transfer these properties to the concept of language itself. But this motivation for structure disconnects it from any *explanatory role* in describing or explaining the understanding of actual speakers. Perhaps some would see this as something of a pyrrhic victory, although others will be heartened by this disavowal of explanatory pretensions.

Still the niggling question persists: why should this strategy be successful? What is it about the speakers' use of L that sustains this decision? Why are the data gathered from speakers' usage of L more or less amenable to certain postulations of structure? Surely, anyone who claims that the decision to construct a structure-discerning semantic theory is correct only if that decision is borne out by the evidence owes us some explanation of why it is or isn't borne out in particular cases. Some account is due of what it is in virtue of which the theorist either does, or does not, get away with his theoretical descriptions. The thought that there might be facts about semantic structure *already there* in the language to be described by the theory runs counter to the attributionist's position.

One promising line suggests itself immediately. Perhaps we can claim that the re-combinable, recurrent and projectible parts and structures of sentences of L are not only described by means of the articulated semantic theory, but are also perceived by the speakers of L themselves? It is hard to see how this response will serve. The attributionist believes that the facts of linguistic meaning, including facts about linguistic structure, are constituted by *description*, and thus they are constituted by the articulated structure of the theory that brings them into sight. Facts about a natural language are answerable to facts about its theoretical description in a theory that meets the normative a priori constraints. The language will be

systematic only in so far as the theory is systematic. So where are we to look for evidence that speakers of L perceive, or grasp, the 'facts' that the sentences they hold true are made up out of semantically relevant expressions under some particular mode of combination? These facts are already a matter of theoretical description, so they cannot be used to bolster our confidence in the theoretical descriptions. Is the correct conclusion that unless speakers are radical interpreters, in the strong sense of being theorists of interpretation, they could not know such facts at all? Or could speakers simply be said to know the semantical structures of their sentences in the reconstructed sense that knowledge of truth conditions can be attributed to them as a consequence of the theory which adequately interprets their behaviour? The trouble with the latter suggestion is that the only behavioural data for interpretation must be describable without appeal to the concepts of theory. So what makes it plausible for the theorist's handling of L to import those particular structural discriminations into it?

It is hard to imagine anything more the attributionist could offer to try to ground his theoretical appeal to structure in the practices of speakers given that he is not describing some antecedently available set of facts about actual speakers' linguistic capacities.

The last resort is to appeal to a prior assumption that speakers learn the meanings of some sentences on the basis of others; in fact, that they have to, if their finite resources confer upon them a potentially infinite capacity for understanding sentences. On this assumption, it is as well to construct a theory that shows how it is possible. But appealing to the course of learning to support a theoretical account of a learnable language is out of bounds to the attributionist. For the attributionist believes that facts about mind and language are delivered by a theoretically adequate re-description of a subject's actual and potential behaviour. So if an appeal can be legitimately made to patterns of learning (described in intentional terms), then the details of this process must be included already in the theoretical description; or if such an appeal runs counter to the attributionist stance they are rightfully ignored. But Davidsonians of this persuasion either eschew what they regard as the non-philosophically

illuminating details of language acquisition ("we are entitled to consider in advance of empirical study what we shall count as knowing a language" Davidson 1984, p7); or are balefully silent on accounts of learning ("Light dawns gradually over the whole" Wittgenstein, On Certainty Sect.141, oft-quoted by McDowell) But even so, there can be no appeal to learning because there is no available set of facts, antecedent to the theory, to do with the stages of meaning-conferring linguistic cognition in speakers. The attributionist cannot take these to be psychological facts about speakers which the theory aims to describe. So this plausible avenue of checking on the actual linguistic capacities of speakers cannot feature on this approach. The attributionist has cut off any retreat to facts at another level of inquiry that might have *supported* his descriptions and to which they could be seen as answerable.

But the attributionist still decides to assume systematic dependence of meaning on form in giving his theoretical descriptions of natural language. Whether this systematicity can be borne out or not, (and if so, how?) is left unexplained by the attributionist. Is there no more to be said?

Grant for the moment the unilluminating nature of such a proposal, I think it is still possible to see why the theorist would want to assume systematicity in the first place. As a project, it is a legitimate a priori undertaking. When the question is: "Given that speakers could learn by rational means the meanings of up to an infinite number of sentences on the basis of exposure to a finite number of sentences, how is it possible?" the finitely axiomatised truth-theory supplies an answer. The theory does not aim to describe the *existing* contours of actual speaker's knowledge, rather, it must construct *plausible* contours for it. The plausibility, or otherwise, however, must remain unexplained. Worse still, I suggest that it remains incomplete since the method of rational reconstruction is not very plausible for syntax, and presumably this is part of a speaker's knowledge of language.

Nevertheless, the interpreter's grip on semantic structure and *his* knowledge of semantic dependency is ensured by this approach. They must be antecedently familiar to anyone who can deploy such a theory to

interpret others. But who interprets the interpreters? As I shall argue later, the omission of any convincing answer, when there are places to look for answers, is a glaring omission in this account.

Before going on let me revert to the question of indeterminacy and ask how much indeterminacy has been removed by the rationally reconstructing theorist's way of motivating semantic structure? The answer is: very little. There are a number of ways one can import semantic structure into a systematic truth-theory. For example one could diverge from Tarski and use the Fregean hierarchy of semantic types when treating noun-phrases and treat singular terms as basic terms, and quantifier phrases as second-order predicates of first-order predicates of terms; or one could treat singular terms as second-order predicates of the predicates they can be combined with to express true or false sentences. This would have the knock-on effect of forcing the quantifier variable to be third-order. Such a formalisation is not inconceivable. (See Wright 1983, p33)

Davidson himself has his own preferences, believing that we must read first-order quantificational logic into interpretable language, and hence into the thoughts of a speaker. This leads to a decision about which logical forms can be assigned the sentences of the language: but by itself it cannot lead to conclusions about which assignments of those logical forms to make to particular sentences.

Nor do these theoretical decisions always enable one to identify which sentence is which in any pattern of dependencies. For example, there could be cases amongst the configurations of the above sort, S_1, \dots, S_n and S , where we could swap the meaning of S with the meaning of one of S_1, \dots, S_n without changing the membership of the smallest set of axioms that will enable us to derive all the corresponding T-sentences, and while still preserving the patterns of dependence and revision throughout. Take as a simplified example a language which consists of just the following three sentences: $A \vee B$, $A \vee C$, $B \vee C$. Any two of these could stand in the pattern of dependencies to the other one. With the same axiom set and the same assigned structures there would still be indeterminacy of meaning within the semantic theory for this language. These cases may be

unlikely once more and more sentences have to be treated by the T-theory. The holistic interdependencies of sentence meanings across the language may make these substitutions unworkable in an adequate theory for the whole language. But it remains to be shown that such cases are impossible to conceive within holistically construed theories.

However, another and more damaging source of indeterminacy looms large. I mean, of course, the indeterminacy of reference.

5.3) The Inscrutability of Reference

Suppose that the semantic values assigned by the axioms to each of the semantic primitives common to S and S_1, \dots, S_n could be systematically changed while preserving the semantic roles of the expressions within whole sentences, so that there could be a different smallest set of axioms from which to derive the T-sentences of S_1, \dots, S_n and S while preserving patterns of dependence and revision amongst them. If this can be done in a number of different ways, then a number of different, arbitrary, competing T-theories can be constructed, compatibly with all available evidence, for the language in question.

How can it be done? The example is by now familiar of a shift in the references assigned to singular terms of the language. Briefly, the idea is that each axiom that assigns an object as reference to a singular term is replaced by one that assigns to that term a displaced object, related to the original by a permutation function θ that maps each object in the universe on to another that is R-related to it, for some R. R could be as arbitrary as "the place-or-object-occupying-the-place-two-inches-to-the-left-of", it does not matter. The point is quite general,

...even if logical form and truth are fixed, acceptable theories may differ with respect to the references they assign to the same words and phrases. (Davidson 1984, p228)

This works in the following way for any permutation θ :

Chapter Five

If we have a satisfactory scheme of reference for a language that speaks of this universe, we can produce another scheme of reference by using the permutation: whenever, on the first scheme, a name refers to an object x , on the second scheme it refers to $0(x)$; whenever, on the first scheme, a predicate refers to (is true of) each thing x such that Fx , on the second scheme it refers to each thing x such that $F0(x)$. Assuming that reference is geared to an appropriate characterisation of a relation like satisfaction in each case, it is easy to see that the truth conditions the second scheme assigns to a sentence will in every case be equivalent to the truth conditions assigned to that sentence by the first scheme. (Ibid., p229)

It is easy to see that the structural properties and logical dependencies amongst sentences would not change, but Davidson shows himself to be departing from Frege here in what he means by a sentence's truth condition. For Frege, if the referents, which are the leading protagonists in a truth-condition, change, then so do the senses of the expressions, just so long as sense determines reference. And if the sense constituents of a sentence change then so does the sense of the whole sentence to which the expressions with those senses belong. But since the sense of a sentence is its truth condition, the truth condition will have changed as well. Davidson clearly departs from Frege here, and perhaps this is where we can point out that he has no use for the notion of sense. Truth conditions for Davidson are at the level of reference, or better, ontology, and not at the level of sense (reference-determination). The radical interpreter tries to fix the truth conditions for a subject's sentences, at first, by pairing them with sentences of his own that share the same truth-values. In so far as he can use his own language's logical relations and standards of truth to constrain the assignments of truth-values to the subject's other sentences, to ensure an overall fit with the subject's pattern of assents and dissents, and to endorse the sentences the subject holds true, he has fixed their truth conditions. The same events and conditions obtain in the world under the permutations; the evidence of assent or dissent in speakers is exactly the same on one scheme of reference or the other. So the determinants of truth conditions remain the same while the references are permuted.

(This conclusion might seem to threaten the Fregean identification I suggested between the truth condition of a sentence and the object of the belief of a speaker who holds it true. But I think not. Davidson holds to the identification, but in doing so, he reveals how indeterminate, downgraded and theory-dependent his notion of belief really is. More on this later.)

These alternative schemes of reference give rise to alternative truth-theories, but they do not by themselves lead to indeterminacy. There is indeterminacy only if there is no way to choose between such theories, no way of picking out the T-theory with the correct scheme of reference for speakers of L. To be able to do this we must find some condition, or set of conditions, that underpins the relation of reference. With such an account one could provide a strong adequacy condition on interpretative theories of truth for L. Reference could serve as a constraint on interpretation.

Alas, the attributionist has denied himself all the materials with which to choose between the different theories that anchor language to the world. The natural place to look is in the mind of the speaker. Intuitively, one wants to ask, which object is he thinking of, x or $0(x)$? But remember that the only materials at our disposal here are the propositional attitudes ascribed to a speaker in the course of interpreting his behaviour. And it has already been claimed that belief and meaning are inextricable and the best grounds for attributing belief to a person rest upon an interpretation of his speech. So any resolving of the indeterminacy must proceed on both fronts at once. An account of reference must solve for belief and meaning simultaneously.

With this in mind we can turn to another intuitive proposal, offered here by Colin McGinn, and this is the idea that causal links between the speaker/thinker and his environment underpin his ability to refer to, or think about, one object as distinct from many others. After all, these are relations the interpreter can take into account from his third-person perspective.

McGinn states Davidson's necessary conditions for something's being the referent of a speaker's singular term, or the object of his thought.

Chapter Five

Heeding the inextricability point, McGinn gives the conditions for both thought and language:

.....an object x can be the referent of a term t in a sentence S_0 uttered by a speaker U only if there is some set $T = \{S_1, \dots, S_n\}$ of sentences containing t such that U is disposed to affirm T and x satisfies the "predicative component" of the majority of $\{S_1, \dots, S_n\}$, for some fairly large ("endless") n . (McGinn, 1977, p524)

The reason why U might be disposed to affirm T is that he has many true beliefs about x , expressed by the sentences in T . But when we look to see which object his beliefs concern, the best we can do is this:

.....an object can be the subject of a belief B_0 only if there is some set of beliefs $S = \{B_1, \dots, B_n\}$ such that x satisfies the "predicative components" of the majority of $\{B_1, \dots, B_n\}$, for some fairly large ("endless") n . (loc. cit.)

McGinn calls this position the satisfactional theory of intentionality. The object the speaker is thinking, or talking about is determined by the speakers beliefs, and assertions, about the properties the object has. The beliefs about its properties pick it out. Of course, there may be false beliefs the speaker holds about that object, but even for that to be the case he must already hold many true beliefs about it (or many beliefs that concur with the interpreter's conception of it) for there to be an object of reference. But since an object is determined by the speaker's beliefs about its properties, if more than one object can satisfy them, then it is indeterminate which object his thought and talk concerns.

McGinn's response to this exploits the relational character of the content of perceptual states, namely, that which perceptual states we are in depends on just which object is involved (in the right way) in the causal process that gives rise to that perceptual state. If one is having a perceptual experience as of seeing an F , and the content of that experience is correctly stated by mentioning an object a which happens to be uniquely F , then one is not having the same perceptual experience in another world where

one is having a perceptual experience as of seeing an *F*, correctly stated by mentioning *b*, an object uniquely *F* in that world, if *a* is not identical with *b*. In each case, the perceptual state necessarily involves its object. So if belief can inherit these characteristics of perception, to have a belief *B* about an object *a*, it is necessary that *a* be involved in some appropriate causal chain that resulted in *B*. As McGinn wants to say, "the fact that *B* is necessarily *about a* is *explained* by the fact that *B* is necessarily caused by *a*" (Ibid., p529)

Now our question is whether the attributionist can take advantage of this response. Certainly, the radical interpreter can take into consideration a subject's relation to his perceptual environment. So an interpreter should be able to include causal relations in the grounds for his attribution of attitudes to the subject. But remember, the only evidence the interpreter has to go on, as far as linguistic reference is concerned, is the linguistic behaviour of subjects with respect to whole utterances. How can one determine from patterns of assent and dissent in the speaker that he does not mean the object or place two inches to the left of the one that he is perceptually in contact with? Sufficient adjustment to the meanings of terms and predicates would secure this as his reference. Secondly, when the interpreter is attributing beliefs to the subject on the basis of their mutual surroundings, the interpreter can only be as determinate as his own beliefs, categories and distinctions admit of. But how discriminating are these? If we are told, they are as discriminating as the causal facts underlying perception allow for, then we have the basis for a more substantial account of mental content, but we give up the attributionist's cherished view that all intentional phenomena have the form and character they do as a matter of interpretation. Causal factors that, in part, explain the interpreter's states of mind and linguistic competence may have to be part of a correct theory of mind and language, but as far as the attributionist is concerned, these factors can only enter the account once they are parts of the interpreter's conceptualised experience. The appeal to causal factors, and the roles they can be given in determining the objects of someone's thought and talk, is a matter of interpretation. There is no going beyond interpretation:

Chapter Five

What a fully informed interpreter could learn about what a speaker means is all there is to learn; the same goes for what the speaker believes. (Davidson 1986, p315)

In the light of this, I disagree with McGinn that interpretation theory:

....is not equipped to explain why it is that the identity and existence conditions of the contents of mental states and of the senses of names are determined by the particular objects whose mention serves to specify those contents or senses. (Op cit., p528)

It is equipped to explain this, but only in its own terms, and not by going beyond the sphere of human experience to point towards causal determinants of the character and content of that experience. Interpretation theory will explain the role of objects in determining the content of someone's attitudes and utterances but only through the interpreter's eyes. Appeal to the external objects is mediated by the interpreter's beliefs about those objects.

One way of trying to get outside the sphere of human experience is to take the barest evidential base and try to reduce the higher level notions to that. This was Quine's way in Word and Object. We now know Davidson's reasons for thinking that that cannot be done. Nevertheless, he does want his theory of interpretation to illuminate the conditions for believing or meaning something, and, in general, making sense:

If we want to illuminate the nature of meaning and belief, therefore, we need to start with something that assumes neither. (Davidson 1986, p315)

Only thus can he show how these notions can be accommodated in the material world. Like Quine he chooses hitherto uninterpreted evidence of a speaker giving his prompted assent to a sentence. The informative part of interpretation theory lies in finding the cause of that assent. Both Quine and Davidson start from the same evidential base. Both are inquiring after the cause of assent, by attempting a backwards inference

from effect to cause. But for Quine, the cause, and hence the interpretation of the sentence, depends on stimulation of the sensory surfaces of speakers. From these meagre materials Quine sticks scrupulously to the impoverished interpretation (or translation) known as the "stimulus meaning" of the sentence.

This lowly view of the cause for speaker's assent is explicitly rejected by Davidson. As I have said, for him, the cause of assent depends on two interlocking factors: what a speaker means by a sentence and what he believes about the world. The radical interpreter must reconstruct these notions giving them application to the uninterpreted speaker's behaviour. Quine's sensory stimulations are no help with this.

....sensory stimulations are indeed part of the causal chain that leads to belief, but cannot, without confusion, be considered to be evidence, or a source of justification, for the stimulated beliefs. (Davidson 1986(b), p317)

Instead, Davidson makes interpretation (the identification of these two factors), "depend on the external events and objects the sentence is interpreted as being about" (loc.cit.) These external events and objects are interpretations going beyond Quine's sensory information and out into the world. But Davidson can only appeal to these because they are interpreted already, they belong to the beliefs of the interpreter and speaker. So Davidson's "causes" are not independent of the interpreted subject's beliefs or meanings:

....we can't in general first identify beliefs and meanings and then ask what caused them. The causality plays an indispensable role in determining the content of what we say and believe. This is a fact we can be led to recognise by taking up, as we have, the interpreter's point of view. (loc.cit)

The causes of the brute behaviour are given through the interpretative sights of the theorist. Davidson can engender a richer result than Quine in interpreting the effect of assent behaviour, because he is not sticking to a brute description of the cause: for this too is interpreted. The causally

relevant events and objects are not independent of the speaker's belief; like the belief itself, they are individuated by the interpreter, who regards them as what causes the subject's belief. The interpreter then charitably ascribes to the speaker the belief about the circumstances those causes engender in him (or, more accurately, 'the belief that includes those events and objects'). The interpreter can only think of them at all in virtue of his beliefs about the circumstances. Thus the events and objects that cause a speaker's belief are presented from the point of view of the interpreter: they are the objects of his beliefs.

Given the interpreter's largely charitable assumptions, he will attribute his own belief to the subject. This happens as follows. The object of the interpreter's belief will be taken to be the cause of the subject's belief, and under the methodology of interpretation the object of the speaker's belief will be the causes of that belief. And most of the time, because of the Principle of Charity, the assumption will be that in the same perceptual circumstances, the same belief will be engendered in the speaker and the interpreter. So as interpreters of what others believe, we infer the content of their beliefs from what causes them. This is how, interpretatively speaking, the possibility of shared belief and communication opens up on the Davidsonian picture. But all of this is secured from within the seamless realm of human experience for Davidson permits no scheme-content distinction. There is no more direct way the mindless world can figure in interpretation. Communication rests on this convergence of causes, but causes which are already objects of belief.

Beliefs, thinks Davidson, can only be identified and supported by their relations to other beliefs. An interpretation which finds a subject right on many matters of fact, identifies the causes of his beliefs with the objects of those beliefs. The interpretation "works" if we can make the identification compatible with the evidence of the subject's behaviour and our expectations for his future action. But the interpretation strictly proceeds from within what the interpreter believes to be the case, and unless there is massive agreement between interpreter and interpreted subject, there will be no sense to be made of the latter.

Chapter Five

Because we are self-interpreting animals, and so subject ourselves to the above points, limits must be placed on the amount of intelligible error we can be subjected to. But limits are placed also on where we may look for more finely individuating conditions of reference for (parts of) our beliefs and sentences. Thus Davidson closes the hermeneutical circle. *Verstehen* does not submit to scientific investigation. Matters of mind and meaning are captured in the normative notions of a higher-level intentional vocabulary that cannot be incorporated into, but must be imposed upon the world of facts as described by the natural sciences. Which, of course, is of no help with the original indeterminacy at all.

To sum up. Indeterminacy prevails where we limit the evidential base as Davidson does to the overt behavioural evidence of speakers in relation to whole utterances; their assent, dissent and utterances. Of course, unlike Quine, Davidson has the further evidence of the attitudes to go on, but these too are inferred (though not uniquely) from observable behaviour. The contents of intentional mental states cannot enjoy tighter connections with the physical conditions they are about, for their significance partly resides in the systematic relations they bear to one another, and only in the light of these relations do they collectively enjoy application to the lower levels of description and explanation at all. Davidson is a coherence theorist of truth, arguing that to be intelligible (ie. interpretable) the majority of our beliefs must be true and therefore correspond to reality. Thus he thinks coherence yields correspondence, but correspondence without confrontation. In this vein, he asserts that there can be no empirical foundation for belief or meaning.

It is this last point, together with the prevailing a priori orthodoxy of Davidson's philosophical psychology, that leaves him with no room to look for a cognitive relation between a thinker and an object of thought or reference. In looking for this cognitive relation, the rational psychologist can only consult the beliefs, desires and other propositional attitudes he has grounds for attributing to the speaker. But the grounds for their attribution include the interpreted utterances made by the thinker, and it is these interpretations of the thinker's words that we sought to render more precise in the first place by looking for a cognitive or causal relation

between the thinker and an object of his thought or reference. An appeal to rational psychology offers the interpreter no way out of this circle.

Yet why should everything in the mental realm be treated as a cognitive state whose content is conceptualised by the subject of experience? Why shouldn't we examine other types of intentional state, such as those involved in the early stages of visual processing, or auditory or tactile sensations, to see if we could find more discriminating object-dependent cognitive relations to identify the objects of our perceptually based thoughts? Perhaps non-conceptual informational states play a crucial part in certain cases in providing the conditions underpinning certain modes of identification of the thought about objects by providing us with information-links to the objects. But postulations of such states, and empirical examination of the relevant perceptual links cannot proceed without paying due attention to the subject's conceptualising equipment for receiving and shaping such informational input at the level of thought. So an empirical investigation of the passage of information through all these levels would depend upon a sophisticated and multi-layered theory of cognition that reached right up to the domain of conceptual thinking.

It is just such a (partly) empirical theory of cognition that Davidson rejects as having nothing to do with the mental. He has argued that the application of propositional attitudes to agents is the criterion of mental life, and that any investigation of phenomena outside this realm, where normative notions no longer have application, would amount to "changing the subject". These empirical proposals would no longer have a bearing on the mental directly.

I think this is wholly wrong and shall argue against it in what follows. But here I have simply been canvassing the range of materials the attributionist has at his disposal to cut down on the indeterminacy of thought and language. And in turning to the the rational grounds for belief and desire attribution that characterise a speaker's psychology, the interpreter finds nothing independent to say about the speaker's way of thinking of objects which might substantiate the linguistic interpreter's appeal to the relation of reference. Rather, a person's ability to think about

a particular object is only as secure as the claim that he can talk determinately about that object to the exclusion of all others. This, in its turn, depends on how he may be linguistically interpreted, and this rests with the interpreter's way of construing the speaker's referential talk. But being in the same position as the speaker, linguistically speaking, the interpreter cannot put any more determinacy into his rendering of others than is available in his own linguistic resources.

There is no way to escape the conceptual scheme of the interpreter and look for unmediated causal relations to connect his language or thought to the world more directly. The collapse of the scheme-content distinction forbids this: "it is absurd to look for a justifying ground for the totality of beliefs, something outside this totality which we can use to test or compare our beliefs." (Davidson, 1986, p314)

For this reason too there can be no suggestion that the speaker's words secure objective reference to non-linguistic particulars in a way that is in principle beyond the ken of the interpreter. Of course, there will be occasions when speaker and interpreter diverge in thought and meaning, holding different beliefs about what is true in their shared present circumstances. But it does not follow that we can thereby impute to the sentence uttered by the speaker more determinacy of meaning or reference than a well placed interpreter could attach to it. Divergence is catered for within interpretation theory, that is why we need to attribute beliefs as well as meaningful utterances to a speaker: belief takes up the slack between what the speaker holds true and what is objectively the case. But divergence does not entail that the speaker's belief or meaning on such occasions falls outside the interpreter's compass. In those circumstances, the objects and events give rise to (cause) different beliefs in the speaker and the interpreter. This shows up when the speaker judges the sentence assented to to be true (false) whereas the interpreter takes it to be false (true). If the interpreter makes a theoretical conjecture that he and the speaker mean the same thing by the sentence, then he has to conclude they have different beliefs. But does this mean that the character of the speaker's belief is determined by some objective features of the events that caused it, quite beyond the interpreter's epistemic

Chapter Five

reach? Beliefs do not have their contents fixed one by one in this isolated way. The correct divergent belief to ascribe to someone here is one that does service in explaining the speaker's reason for asserting (denying) the sentence, from the interpreter's point of view. The interpreter cannot just defer to the causal conditions in giving the content of a speaker's beliefs. Instead he must cast around to find a belief, belonging to a set of beliefs which forms part of an interpretation that make the speaker's decisions, and thoughts and utterances intelligible to the interpreter. So even in the case of a single belief, sponsored by the ambient conditions, that belief must be individuated and tested holistically: "all that counts as evidence or justification for a belief must come from the totality of belief to which it belongs" (Ibid., p319) - a totality that makes best sense of the speaker's speech and conduct.

So after all these points have been heeded, what does the notion of reference amount to for the Davidsonian? Whatever role it has, it is far from substantive since:

It plays no essential role in explaining the relation between language and reality. (Davidson 1984, p225)

Once we have gone this far, what useful function can reference serve? Why don't we just discard it altogether? Some of Davidson's headier claims seems to suggest that we can:

We don't need the concept of reference, neither do we need reference itself, whatever that may be. (ibid., p224)

But this is too strong, and all Davidson means by it is that:

..if there is one way of assigning entities to expressions....that yields acceptable results with respect to the truth conditions of sentences, there will be endless ways that do as well. There is no reason, then, to call any of these semantical relations 'reference' or 'satisfaction'. (Davidson 1984, p224)

It is too strong to say that we don't need reference at all because even when stripped of its usual extra-linguistic function, it must be retained to

preserve the compositional insight that enables the interpreter to assign truth conditions to every indicative sentence of a speaker's language. The thought here is that to assign a given sentence a particular truth condition, and to have it depend holistically on the assignments of truth conditions to all the others, one must give the sentences an articulation which exhibits the expressions that recur in other sentences, and assign the same semantic roles to these expressions in other sentences. That is to say, to interpret a piece of linguistic behaviour as showing that a speaker assigns a truth condition p to a given sentence S , one must interpret his linguistic behaviour with respect to other sentences structurally related to S : this can only be done by providing systematic descriptions of the inner structures of his sentences in terms of names, predicates, quantifiers, connectives and functors, and giving each of these a semantic role that contributes to (partly determines) the truth-values of sentences containing it. This is where reference enters the picture, for the compositional insight is cashed out for the interpreter by a truth theory that shows how a sentence's truth-value is determined by its structure and the reference of its constituents. Truth theory shows what it is to treat a sentence as having a certain truth condition by showing how the referential aspects of that sentence have the semantic function of determining the truth-value of this and other sentences in which they occur. Moreover, it is by fixing reference for a finite number of basic vocabulary items and fixing the truth-affecting properties for finitely many constructions that the theory recursively explains the truth conditions for infinitely many sentences. So how could such a theory serve to interpret a language if we denied it the truth-affecting nature of the subsentential expressions?

Davidson's solution is to preserve the theory-internal function for reference assignments while denying that they have any empirical import. This squares well with the idea that we certify a truth theory empirically by testing (some of) its consequences. But it also means that we cannot test the truth conditions a T-sentence assigns to a sentence by assessing whether the right referents are given the leading roles in determining its truth-value. This would be to suppose that we have some non-truth-theoretical way of determining the references of singular terms and

predicates: but this is just what Davidson denies us. According to him, there is no empirical check on reference, only on truth; so reference assignment is not an empirical constraint on the correct ascription of truth conditions, it is merely a theoretical means of achieving this.

This strategy calls for a distinction between explanation *within* a theory and explanation *of* a theory. Explanations *within* a theory invoke the machinery of reference and satisfaction in the axioms of the theory, assigning these properties to the structurally revealed parts of sentences. From the axioms, together with simple inference rules, we can derive T-sentences of the form: "S" is true-in-L iff p for all indicative sentences of the language L. The derivational structures of the T-theorems reveal the logical forms assigned to the object language sentences; and this is how the compositional insight is promoted.

On the other hand, explanations *of* the theory have to relate theoretical concepts to the uninterpreted facts at the lower, physical level of description, and the suggestion here is that these points of contact can be described without mention of sub-sentential semantic structures or the sub-sentential reference assignments. The theory makes contact with its empirical base at the level of its output theorems, and not at the level of its axioms. It is always related to that evidence in the context of a rational explanation of a speaker's linguistic behaviour (overall theory of interpretation), and the confirming fit with the T-sentences is always holistic.

The distinction between explanations *of* a theory and explanations *in* a theory enable Davidson to propose a molecular truth theory as part of a holistic theory of meaning. It is only when the molecular truth theory is construed in terms of the empirical evidence and its role in interpretation that it functions as a holistic theory of meaning. This point also shows that it is not in every case necessary to relate the theoretical concepts used by the interpreter to the empirical base for interpretation. Some merely serve to oil the internal wheels of the theory. Let us call these the *purely theoretical concepts*. So while the concepts of truth, reference and satisfaction are all theoretical concepts, in the sense of belonging to the mechanics of truth theory, truth alone corresponds for Davidson to a pre-

Chapter Five

theoretic notion; reference and satisfaction are purely theoretical concepts. This much is unequivocal in Davidson:

I suggest that words, word meanings, reference and satisfaction are posits we need to implement a theory of truth (Davidson 1984, p 222)

What we need to know is how extensively Davidson applies this pure theoretical status to other semantical notions at, or below, the level of the sentence. For instance, is the logical form of a sentence a purely theoretical notion? Is it similarly empirically indeterminate, and only necessary for specifying the truth conditions of all sentences? Davidson's answer is far from clear. I suggest that it depends on what he means by saying that the form of the theory can be *reconciled* with the nature of its evidence. On the one hand we have it that:

The articulation of sentences into singular terms, quantifiers, predicates, connectives.....must be treated as so much theoretical construction, to be tested only by its success in predicting the truth conditions of sentences. (Davidson 1984, p74)

which leads Davidson to declare that it is one of the chief merits of this approach that:

it touches the observable only when it comes to sentences, [and] seems to *leave the analysis of internal structure simply up for grabs* (Ibid. emphasis mine)

On the other hand, Davidson declares that:

A satisfactory theory for interpreting the utterances of a language, our own included, will reveal *significant semantic structure* (Davidson 1984, p130 emphasis mine)

He also tells us that these 'structurally-revealing' theories have done their job once they have delivered interpretively adequate T-sentences,

and that these make "no mention" of the internal apparatus of the truth theory. This will take a little effort to unravel.

In what follows I will bring out two objections to the Davidsonian programme. Both of them address the fact that for Davidson the notion of meaning is an artefact of truth theory, and as such has a theoretical status which demands a certain theoretical character in someone's knowledge of meaning. The objections are (1) that his theory of language does not describe as much as speakers ordinarily know, and (2) that it credits them with more than they ordinarily know. Now both of these objections depend on the assumptions that a theory of meaning should be a theory of understanding, and that there are facts about a speaker's knowledge of meaning against which to test our accounts of meaning. These are highly controversial assumptions and may mean something quite different in the Davidsonian setting. So we should do well to proceed carefully. But just to signal the strategy of the argument in advance, should it prove philosophically cogent to substantiate the claims about a speaker's linguistic understanding the claim against the Davidsonian programme will be that it cannot serve as a theory of meaning because it fails to recover what speakers know of the meanings in their language, and it fails as a theory of understanding because it credits them with knowledge they do not have. So the first step in this line of reasoning is to establish what the Davidsonian can recover of meaning for any given language, while at the same time trying to ascertain what kind of understanding he credits to speakers of the language in question. In particular we should ask does meaning reside within the theory or in the linguistic evidence, and do speakers of the language have access to the theory or just the linguistic evidence?

5.4) Logical Form as Significant Semantic Structure?

Davidson's talk of reconciliation between theory and evidence (molecular/holistic; reference-specifying/referentially-inscrutable; sub-

Chapter Five

sententially-structured/ sententially tested) is premised on the idea that the formal theory of truth "imposes a complex structure on sentences" as revealed in the proofs of the corresponding T-sentences, while the evidence for the T-sentences, and hence for the theory as a whole, consists "entirely of facts about the behaviour and the attitudes of speakers in relations to sentences". So how do these facts confirm the *significance* of the structure shown in the proofs? Is there some way, perhaps obliquely, in which evidence of this sort harbours structure *revealed* by the correct theory? Not as far as I can see. Though, in his early papers, Davidson was given to saying that:

Empirical power in such a theory depends on success in recovering the structure of a very complicated ability - the ability to speak and understand a language (Davidson 1984, p25)

Nowadays, it is much less clear that he would have any truck with common structure in the abilities of all English speakers, or in the ability of a single speaker across time. Instead, a speaker's linguistic ability is seen as a method for converging on interpretations of his own and others words from time to time. But:

There is no more chance of regularizing or teaching this process, than regularising or teaching the process of correcting new theories to cope with new data in any field - for that is what this process involves. (Davidson 1986, p446)

On this latest view there is no common method of interpretation across speakers, or single method in a speaker across time; and since languages are identified via theories of interpretation, this means that is no such thing as a common language, or even the language of a single speaker. But what of the case of syntax, given that syntactic criteria partially identify languages? Are these facts similarly a matter of varied, idiosyncratic conjecture? No, Chomsky would argue; but he trades on something like Davidson's despairing conclusion to argue that in large part syntactic knowledge must be innately specified. Davidson is simply looking in the wrong place for an account of the underlying regularities of

Chapter Five

our language. Chomsky would agree that syntax cannot be taught - but he has shown that it can be regularised. I shall come back to this too. Thus, while having reservations himself about the idea of an external, and only extensionally characterised public language, Chomsky would resist Davidson's latest eliminativist conclusion that:

....there is no such things as a language, not if language is anything like what many philosophers and linguists have supposed. There is therefore no such thing to be learned, mastered, or born with. (Ibid, p446)

This eliminativism about language is not easy to reconcile with many of Davidson's past remarks, but perhaps it is the inevitable outcome of what I have been calling his Semantical Minimalism about meaning. I shall restrict myself to an inquiry into this minimalist view of meaning, rather than look at the newly arrived at eliminativism. But this is still apposite since meanings, or rather interpretations, do survive in the new and somewhat dispiriting picture.

Certainly, Davidson's earlier remarks are more encouraging; but the evidence of the linguistic ability he mentions has to come from evidence of behaviour and/or evidence for psychological ascription, and neither of these seem to serve his purpose. To begin with, there are only two ways of construing talk of behaviour: it is either a matter of the hard physical facts, or it is a matter of the intentional behaviour of the agent. Construed in the first way, it is hard to see how the physical properties of speech events could display the semantic structures of sentences since, as Davidson himself admits:

The relations between these semantically tractable patterns and the surface grammar of sentences may.....be very complicated. (Ibid., p151)

One need only recall Davidson's handling of adverbs in action sentences, where he treats sentences as involving implicit quantification over events, to recognise how revisionary of the surface structure (word

Chapter Five

strings?) these logical forms are. Obviously, the extra argument place cannot show up in the *physical* properties of the utterance-event.

Should we look for the 'recoverable structure' of a speaker's linguistic abilities in his intentional behaviour then; that is, in his linguistic actions? Well, since we are not looking for signs of logical form in the external behavioural surfaces of an action, perhaps we should look for a deeper source of evidence at the linguistic attitudes in the mind 'behind' (embodied in) these utterances. However, the way things have been set up in interpretation theory there is no hope of examining the mind with a view to confirming particular sentence structures since they themselves are used to give a more precise form to mental descriptions:

...attributions of attitudes, at least where subtlety is required, demand a theory that must rest on much the same evidence as interpretation [of linguistic behaviour]. (Ibid., p134 bracketing mine)

So in neither the hard physical facts of behaviour, nor the intentional character of the mind do we find a recess in which to locate the semantic structures brought to light by the theory. So what are we to make of Davidson's talk of "significant semantic structure" ? He explicitly denies that semantic structure can be regarded, as Chomsky regards syntactic structure, as reflecting "the internalized grammars" of speakers of the language. He even goes so far as to deny that we need postulate the differences in syntactic deep structure Chomsky points to in pairs of superficially similar sentences like:

- (1) I persuaded John to leave
- (2) I expected John to leave

which show meaning differences in the active and passive complements of "persuade" but not in the active and passive complements of "expect". (Eg: "I persuaded John to be examined by a specialist"/ "I persuaded a specialist to examine John" ; cf. "I expected John to be examined by a specialist"/ "I expected a specialist to examine John")

Chapter Five

Instead of showing their differences as Chomsky does (1981, p100) in terms of what he regards as the different syntactic descriptions a speaker's internalised grammar assigns to these sentences:

(1') I persuaded John [PRO to leave]

(2') I expected [John to leave]

where "PRO" is the empty Pronoun Subject and "John" in (1') is the main clause object and not the subordinate clause subject, thus showing a grammatical similarity between (1) and sentences like:

(3) I decided [PRO to leave]

(4) The president isn't sure [whether PRO to vote for himself]

Davidson supposes that the difference between (1) and (2) is one of semantic structure shown by noting that "John" can be replaced by any co-referring expression in (1), *salva veritate*, but not in (2). This would make the semantic role of "John" different in (1) and (2) and this would have to be recorded in truth theory. I doubt whether these intuitions are very reliable, but notice anyway that in Davidson's attempt to motivate an alternative to deep structure in terms of the semantic structure of truth theory, he does not explain the difference in semantic structure in terms of a speaker's tacit knowledge, or the "intrinsic competence of an idealised speaker". Instead, he explains it by claiming that "the evidence suggests that a theory that assigns different structures to (1) and (2) may be simpler than one that does not" (Davidson 1984, p64). But unfortunately, Davidson's suggestion and his simplicity assumption for *this* case will not provide the extensive generalisations about structure provided by the postulation of PRO and the consequent structural arrangements offered by Chomsky. It turns out that the simplest theory that deals with all the data needs far more than the categories of semantic structure, and that it is far from simple (or rationally inductive) at all. (I shall elaborate on this when we come to discuss syntax below.)

Chapter Five

This claim of Davidson's that truth theory can deliver propositional forms from word strings is wholly untried; but equally, from the empirical evidence, it is wholly unlikely. So he could only ever be claiming to be describing some *part* of the structure of a speaker's linguistic ability, for the devices of the truth-theoretic interpretation could be applied to a language only after it had been regimented by something like a generative grammar; and it would be the empirical claim of syntactic theory that it describes aspects of the cognitive psychological structure of a speaker's linguistic ability. But since this is not what Davidson has in mind for semantic theory, there would have to be both an *a priori*, attributed structure and an underlying cognitive psychological structure to someone's linguistic ability. But if there are any significant relations between syntax and semantics we shall soon confront the thorny problem of how this interface is to be described.

That said, what is the justification for saying that the structure of semantic theory reveals the significant structure of the language? So far, then, all we can find reason to say is that semantic structures reside entirely on the side of the theory, and that the theory is imposed on the evidence. But we have still to say what it means "to reconcile the need for a semantically articulated structure with a theory testable only at the sentential level" (Davidson 1984, p137) - where, presumably, at this level, the sentences do not display the complex logical forms they are given in the theory. Perhaps this can be done by a theoretical argument that claims that it is only by bringing the theory to bear on the evidence that one can say that someone's sentences have a meaning (interpretation) at all. The justification would be that meaning is a matter of a sentence's having truth conditions, and imposing a truth theory brings meaning to light because it is only in this way that one can justify talk of sentences possessing truth conditions.

Now we seem to be getting somewhere; but the appropriate cost will be that a speaker's language and the meaning of his sentences will be a matter of theoretical construction; notions externally attributed to him on the basis of his behaviour. Semantic structure is in the eye of the theorist, who must impose it on linguistic behaviour in order to produce

Chapter Five

interpretations. So it is a theoretical presupposition of the interpreter that he will find semantic structure in his subject's sentences: it is this imposition of form on the evidence that makes it possible to treat a subject's sentences as having truth conditions. This looks like the right thing for the attributionist to say. And as we know Davidson even tells us what the presupposed forms are:

A satisfactory theory cannot depart much, it seems, from standard quantificational structures or their usual semantics. (Ibid., p151)

This is a theoretical decision to read into the object language the structures of quantificational theory employed in the metalanguage:

The result of applying the formal constraints is, then, to fit the object language as a whole to the procrustean bed of quantification theory. (loc. cit.)

Note, however, that it is not obvious here whether this itself is the process of regimenting the language for interpretation, or whether that can only be done once the theory of syntax has extracted linguistic forms from the primitive data. Either way, it is little wonder that the interpreter 'finds' certain significant logical forms in the linguistic material he interprets: he simply reads them into the object language; and little wonder too that the differences in the structure of truth theories for given languages will be minimal, with the major differences consisting in the content of their axioms. However, this imposition of structure does not by itself entail, as Davidson seems to imagine, the stronger conclusion that:

The identification of the semantic features of a sentence will then be essentially invariant: correct theories will agree on the whole about the quantificational structure to be assigned *to a given sentence*. (loc.cit. Italics mine.)

While I accept that, on this picture, there is no difference between theories in the forms that they can impose on the data, there can still be differences in their actual impositions of those forms. For Davidson, all

Chapter Five

equally acceptable theories for a given language share the same set of forms to impose, but different theories could assign semantic structures differently to the individual sentences of the object language. And in that event, if there was no empirical basis on which to choose between the different acceptable theories, there would still be an indeterminacy of semantic structure within the object language sentences themselves.

Admittedly, this would be indeterminacy within principled limits: namely, those set by the format of a Tarski-type truth theory. But it would be an indeterminacy of logical form nevertheless. Davidson's unargued for claim that different empirically equivalent truth theories are in agreement (nearly enough) about the roles they assign to sentences in the language - roles based on the structural relations they are deemed to enter into - (see Davidson 1984, p225) is not established by the more restrained claim that:

...indeterminacy is automatically put under greater control if one insists, as I do, on a Tarski-style theory of truth as the basis for an acceptable [interpretation theory]. (Davidson 1984, p228)

Davidson has not banished, but merely cut down the options for widespread indeterminacy of structure. Although it still remains a distinct possibility that:

...logical form may be indeterminate: two satisfactory theories may differ in what they count as singular terms or quantifiers or predicates, or even with respect to the underlying logic. (loc.cit.)

Let us look at the strategy of containment. It depends on what Davidson calls:

...a principle of reverse charity that judges a theory better the more of its own resources it reads into the language for which it is a theory. (Ibid., p228-9)

But Davidson acknowledges the lack of complete freedom here. Not just anything "can be read into the object language simply by assuming it

to be in the metalanguage". For instance, Davidson would resist assumptions that we are dealing with a modal language just because we can make appeal to metalinguistic modal operators. But what substance can he give to a principled distinction? There are two questions here. One has to do with what semantical resources there should be in the theory; and the other has to do with how the theoretical devices apply to actual linguistic practice. On the view I favour, attempts to answer the latter question provide a means of answering the former. But as we shall see, for Davidson the priority is reversed.

Whether Davidson's application of the reverse charity principle is correct in the case of particular languages would require detailed considerations of how successful he is in using the structures of first-order extensional logic to deal with the semantic phenomena that seem to call for alternative logical forms. These will include phenomena that call upon intensional or modal logics, such as generics, conditionals and modal terms, and intentional verbs. Such a study would also examine the proposed refinements to the Davidson framework to deal with cases of vagueness, tense, discourse reference, demonstratives and indexicals. Full discussion of these issues is obviously beyond the scope of this work, but we can still raise the important question of why Davidson assumes that some and not other theoretical characterisations of a language are *appropriate* ways to characterise language. Davidson reads this as the first of our two questions - what theoretical devices should the theory make use of - and he hopes to subordinate the second - how do these notions relate to linguistic practice - to it. The reasons for this will emerge, but let me stress once more that without an answer to the second question we shall have no idea why the strategy of attributing quantificational structures to object language sentences is successful in capturing regularities in language use, patterns of speaker's inference, and projection to new cases.

Surely it is quite natural to think that when a truth theory provides a successful interpretation of a given language there must be some evidential feature of a speaker's linguistic practice, or some pattern or structure in his ability to use the language for the logical forms supplied by

the theory to pick out? But it is just this thought that an attributionist like Davidson denies to us. Is it then correct, in the absence of any significant structure in the evidence, to resort to the idea of a forced fit with the forms of the truth theory? That move would seem to be hopeless because it threatens to allow back into the picture the postulation of alternative logical forms, with the resultant indeterminacy. But what other move can the Davidsonian make? The pattern discerned in a sentence is given by the theory and is not there in the evidence independently; but at the same time, Davidson believes that the evidence of that sentence's use does not permit the imposition of the several different patterns which different kinds of semantic theory would each see it as exemplifying. Clearly, Davidson urgently needs some way to motivate the idea of an *appropriate fit* if he is going to argue not just for finite axiomatisation and structure-revealing theories, but for a certain form of axiomatisation and a certain range of logical forms. However, what he resorts to, and what blocks the threatened return of alternatives which the idea of a forced fit seemed to enjoin, is the provision, on a case by case basis, of auxiliary arguments at the meta-level for why we should not make use of alternative forms of semantic theory. In the modal case, he offers several arguments that we should not employ an object language and metalanguage that differ in expressive power (I shall not discuss these arguments here since I am concerned in the main with Davidson's strategy of argument). In the case of adding more structure to sentences on the right-hand sides of T-sentences than we commonly assume for the sentence mentioned on the left, Davidson must distinguish between trivial cases where we just append a tautology to the right-hand side of each T-sentence - such as: (T) "Snow is white" is true-in-L iff snow is white and $2+2=4$. - and cases where the extra structure uncovers a *genuine logical form*, as in the case of action sentences. Perhaps methodological arguments from simplicity will help to rule out the trivial cases, but it will not help with cases where different, equally simple theories provide alternative, empirically adequate, analyses of the structures of sentences. Davidson's position seems ill-equipped to cope with such cases, for he has no uninterpreted source of evidence against which to measure the two theories. He would

Chapter Five

have to resort to a priori grounds to decide the matter, or to rule out the possibility of any such case. But as Dummett says:

...the idea that something may be complex and yet be exhibited as complex in different, equally good, ways is not in itself absurd (Dummett 1981, p279)

We may harbour the intuition that the sentence has one complex structure and not another. Permutations of structure-assignments to the same set of sentences, if compatible with the overall evidence (and we have yet to be shown why they are not) may provide cases where two theories assign different complex structures to the same sentence. Arguably, there would be nothing to choose between such theories: but this would be to concede a point about indeterminacy and forfeit an invariance in the precise roles given to individual sentences by any adequate theory. Perhaps this much Davidson cannot "put under greater control".

The point of all this is to establish is that theoretical possibilities of rival accounts of the semantic character of a language are not closed-off in advance by Davidson's heterogeneous arguments against alternative approaches. Unlike those who think there are facts of the matter about speaker's knowledge or a linguistic practice that explain the linguistic evidence we are trying to describe, Davidson cannot appeal to the linguistic character of the language to support the type of theory he favours, and thus he cannot ignore the recherche alternatives that can be put up as alternative ways of fitting the evidence. He offers a theoretical view in contrast to other theoretical views: so he must pit his theory against alternative views rather than seek direct confirmation that he is capturing the linguistic facts about the language. For theory constitutes these facts according to the attributionist.

Perhaps it will be objected that this is going too far. After all, the empirical evidence does not permit the theorist of meaning complete freedom either. Surely, he is constrained by the patterns in a speaker's linguistic behaviour to fix the meaning of an expression the same way each time it is used. Well strictly speaking, this depends on how much

regimentation of the language takes place prior to interpretation. For instance, can we assume that *word* -strings are available in uninterpreted linguistic behaviour rather than mere emissions of noise? Perhaps the thought is that the interpreter will be empirically bound to assume (subject to special considerations) that an item of the same phonetic type must be interpreted everywhere in the same way. Were he not to do so, he could not assume that expressions made a uniform and stable contribution to the meaning (truth conditions) of sentences in which they occurred. And yet this is vital since it is only by guaranteeing this for each significant non-indexical expression of the language (the indexicals having been cordoned-off somehow) that the interpreter is able to work out the meaning (truth conditions) of every (non-indexical) sentence in the language as a function of the meaning of its parts. However, this supposedly *empirically* well-attested assumption is really a matter of theoretical decision too. For the decision to identify a part of two or more sentences as the recurrence of the *same* word is a decision to recognise a common semantic feature of those sentences; and this in turn depends on a semantic theory which articulates the inner structure of those sentences and isolates a part within each of them that has that unique semantic role. Furthermore, Davidson believes that this theoretical insight into parts and wholes must come from a holistic theory of meaning. The key thought here is that because we cannot fix the truth conditions of a sentence in isolation, we can only fix it in the context of a theory that fixes the truth conditions for every sentence. This is a highly contentious claim and Davidson does little to substantiate it but we can sketch, however unsatisfactorily, how it is supposed to emerge for him.

The truth conditions the interpreter assigns to a single sentences must depend on the semantically significant parts it has; but as we now know this is the result of a theoretical decision. From the theorist's point of view it is by imposing an internal structure on a sentence that it is revealed as having sub-sentential parts whose semantic roles determine its truth conditions. But now this assumption will be confirmed or revised in the light of similar theoretical decisions made about the inner structures assigned to other sentences, and the truth-determining roles of *their* parts.

Chapter Five

The thought here is that fixing the semantic role for any one expression will depend on how one decides to fix the semantic roles of each of the other expressions that accompany it in the different sentences in which it occurs; for it must combine with the semantic roles of different combinations of these expressions to determine the truth-values of sentences in which it occurs; just as the semantic roles of each of these expressions must likewise conspire with the roles assigned to *their* accompanying expressions to determine the truth values for all sentences in which *they* occur. It is only when we have a satisfactory theory that determines the correct truth-values of every sentence on the basis of its theoretical imposed structure and theoretically described parts, that we can certify the identification of the common semantic features of sentences that share the same term or predicate. The semantic role of any subsentential expression will be displayed in truth theory by how the axiom for it functions in the derivations of T-theorems for sentences containing it; and of course this will depend on the supporting role of the other axioms. The correct assignment of truth conditions to any individual sentence will be one that fits into an empirically confirmable and systematic account of how the truth conditions of the sentences of that language are determined; and the correct assignments of semantic features to words will be those assignments that serve to deliver the correct assignments of truth conditions. So to extract the semantic properties of a word from linguistic usage demands nothing short of a theory that has application to the language as a whole. Davidson's takes it that his holistic conception of meaning is a straightforward consequence of this theoretical route to meaning:

Words have no function save as they play a role in sentences: their semantic features are abstracted from the semantic features of sentences, just as the semantic features of sentences are abstracted from their part in helping people achieve goals or realize intentions. (Davidson 1984, p220)

Now while most theorists of meaning would assume with Davidson that, since a language consists of an infinity of actual or potentially utterable

Chapter Five

sentences built up from a finite stock of vocabulary items and constructions, we must see "the semantic features of this potential infinity of sentences as owed to the semantic features of the items of the finite vocabulary" they would not necessarily concur in thinking that the fixing of the semantic features of the basic vocabulary items depends on the language as a whole. That is, they would not share Davidson's view that because the meanings of words depend on the meanings of sentences in which they occur, and how *they* are composed of words whose meanings are given in terms of stable contributions to yet other sentences, that to give the meaning of any word or sentence we should have to account for the meaning of every sentence; ie. that there is no partial ordering of sentences in the language in respect of their dependencies of meaning. Opponents would argue that we have yet to be shown a formally satisfying argument that adoption of the context principle entails a semantic holism where the meaning of a sentence is a function of the total assignment of meaning to every sentence. Michael Dummett (1978), Neil Tennant (1987), Christopher Peacocke (1986a), and Alan Weir (1985) have all offered arguments that it does not. The alternative is semantic molecularism. Molecularism is the thesis that the meaning of a sentence is compounded out of its own constituents and can be given independently of (some) sentences in the language not involving those constituents (although the molecularist must, it seems, permit 'local holisms' in the account of linguistic meaning, thus frustrating any general thesis that sentence meaning is always basic in a theory of language). All the molecularist need acknowledge by way of semantic dependencies is that a grasp of the meaning of any sentence will "depend upon a mastery of some fragment of the language, a fragment which may, in some cases, be quite extensive." (Dummett 1978, p304). I cannot enter into these arguments here (though I will come back to the issue of holism versus molecularity below), but it is worth pointing out that Davidson's view that the semantic features of words are abstracted from the semantic features of sentences does not automatically guarantee semantic holism.

A more promising motivation for Davidson's holistic conclusions about meaning comes from his evidential holism where we have to work

Chapter Five

back to theoretical posits of words and sentences from the gradual application of a formal theory to more and more evidence of the use of the language as a whole. The point is neatly expressed by Gareth Evans. Suppose that a theory treats some expression *e* as *W*, where "*W*" is some property of taking a particular semantic value or being structured, then:

In so far as empirical considerations bear upon the correctness of the claim that an expression is *W*, they bear globally upon the theory according to which it is *W*, and cannot be brought into any more direct relation with that feature of the internal constitution of the theory in virtue of which it may be said to be treating *e* as *W*. (Evans 1976, p200-1)

And even then empirical support for a truth theory will be a matter of how well its linguistic interpretations service the project of interpreting the behaviour of the agent as a whole. The upshot is that the evidence from a speaker's linguistic behaviour cannot support particular meaning-assignments or specifications of structure unless it supports them all. Meaning-specifications for any word or sentence are dependent on all other assignments of meaning made by the theory.

But now we could raise a second objection. Why should we suppose that the only evidence for linguistic meaning concerns the potential infinity of points of contact between theorems of the theory and actual or potential uses of sentences, where this is conceived, at first, in purely behavioural terms? Could we not seek evidence of speaker's *knowledge* of word-meaning or sentence-meaning, as manifested in their linguistic practice as it relates to that word or that sentence? It is just this discriminatory evidence from linguistic practice which Davidson denies us. But the denial rests upon his attributionist picture of the mind of the speaker according to which linguistic and psychological facts depend upon the resources of the interpreter and what theoretical constructions he can put upon someone's behaviour. The constructions imposed on the basis of behaviour are an interlocking set of beliefs, desires and meanings, but we can, with some justification accuse Davidson of neglecting the mental state of a speaker's *knowledge of meaning*, and the neglecting the evidence of *patterns of linguistic use* in which that knowledge is typically

behaviourally manifested. Both of these may be more robust than Davidson imagines. The trouble is that, for Davidson, these patterns of use - let us think of them as behavioural dispositions - are picked out in terms of the flimsy web of intentional notions that overlays behaviour. Conceiving of a pattern of language use in this way, as constituted by an act of interpretation, there is no reason to pay special heed to them: they will be constantly open to revision in the shifting sands of interpretation theory. But if we think of linguistic dispositions of speakers as underpinned by certain cognitive psychological states, which partly identify those dispositions, then it is possible to conceive of speakers as possessing stable mental structures which function as available place-holders for encoding recurrent features of linguistic experience. Of course the character of a disposition *can* change but there is some reason to expect a more stable set up, and to find more continuities in language use, than the interpreter is used to. This makes the regularities in linguistic behaviour less remarkable or mysterious. However, these remarks are highly programmatic, and the alternative conception to which they allude has to be developed before this objection can stick. It will gradually unfold towards the end of this chapter and in the next chapter. But the main point to grasp at this stage is simply that it is a methodological consequence - and hence for Davidson constitutive - of the interpreter's stance that there is *no* reason to suppose that there are stable states of knowledge less given to revision in the light of non-massive changes in a speaker's belief system; there is just no reason to expect any part of that intentional framework to be more stable than any other. So if there are evidentially well-supported generalisations to be had about these knowledge-involving patterns of language use, the Davidsonian with the resources of the radical interpreter will simply miss them.

Instead of speaker's knowledge of meaning, Davidson talks about understanding a language, and a body of knowledge that would suffice for this. But here he is speaking about the interpreter and providing a theoretical account of the knowledge one could attribute to *him*. Davidson is not accounting for the knowledge a speaker possesses, nor is he really accounting for the knowledge an interpreter possesses, save in that he is

talking about the knowledge attributed to the interpreter by the theorist. This view of linguistic understanding will concern us much in what follows. But first we need to know more about his conception of individual meanings; for a proper understanding of this will help us to see why he resists the thoughts about speaker's knowledge of meaning just expressed. Extremely roughly, then, the underlying conception will contain the following strands. Meaning in a language isn't rule-governed in that meaning does not determine use; meaning is abstracted from use in that it is plotted by a theory that interprets a vast range of linguistic and non-linguistic evidence. Just as in the attributionist conception of the mental, there is a certain Wittgensteinian flavour to these ideas. But where Wittgenstein and Davidson part company is in Wittgenstein's very important insight that there is a notion of meaning (grasping a rule) which is not a matter of interpretation (Philosophical Investigations, §201): "Every sign is capable of interpretation; but the *meaning* mustn't be capable of interpretation. It is the last interpretation." (Blue Book, p34). What Wittgenstein's thinking reinforces so powerfully is the idea that so far as *the speaker* is concerned an interpretation will not help him to determine how he should go on extending a pattern in future cases of language use.

Leaving that thought aside for the moment, the next thing to do is to get a clearer idea of Davidson's conception of individual sentence meaning and from here try to find a home for significant semantic structure. In the end these two tasks will be one.

None of the foregoing remarks are meant to suggest that an interpretation theory is entirely unanchored to the empirical facts. It simply means that the empirical facts that anchor the theory at the level of its theorems are very sparse indeed. So is what we find there enough to constitute meaning? If we do not acknowledge the presence of meanings at this level we will rob the concept of meaning of all empirical import, treating it like reference, as a notion purely internal to the theory of truth. This would be disastrous since without empirical conditions of application any theoretically sound assignment of linguistic meaning could be made to accord with the evidence, which simply means that we could not speak

of a *correct* assignment of meaning at all. Meaning would be rendered wholly indeterminate and the attributionist's motivation to accommodate facts about minds and meanings in the material world would have been undercut. This is not an option for Davidson.

Now any empirical content the concept of meaning has must go via the T-sentences, since it is T-sentences that relate to the evidence and nothing else. But do they contain enough information to amount to specifications of meanings? At first it seems not; for remember that these testable T-sentences "make no mention of the machinery" of quantifiers, terms, predicates, connectives, used in the attributed logical forms that take us from descriptions that do not interpret utterances to those that do. On the other hand, Davidson says that "a T-sentence gives truth conditions" (Davidson 1984, p138). But now what sort of conception of truth conditions is this? They make no use of the concept of reference, and make no mention of the constituents of sentences on their left-hand sides. So there is no reason to ascribe a logical form to the sentences they contain: they merely present pairs of sentences of the object language and metalanguage that agree in truth-value. That is why we can test the output of a truth theory without knowing the meanings of the sentences already. But now it will be easy to agree with Davidson that "a T-sentence does not give the meaning of the sentence it concerns" (*loc.cit.*) However, in the light of this and the previously quoted remark, Davidson's claim to uphold a truth conditional conception of meaning now appears highly problematical. For how can these referentially-insensitive truth conditions which make no mention of a sentence's constituents possibly constitute a sentence's meaning? Obviously they cannot; but since they are without logical form they cannot really constitute a sentence's truth conditions either. So what Davidson needs to do, if he is to retain the thought that T-sentences specify the truth conditions (meanings) of object language sentences is to give the notion of logical form application to their left-hand sides. But logical form is a notion derived from the internal working of the theory, it has no empirical basis in the evidence so why should it be included in meaning? The puzzle is why meaning should be aligned with the theory on matters of structure, but aligned with the evidence on

matters of reference. Neither reference nor semantic structure have a basis in the empirical evidence, and since reference is not an ingredient of meaning, or even a constraint upon it, why should we suppose structure is either? Like reference it is just a posit needed to implement the truth theory. One bad answer here would be that Davidson believes, somewhat dubiously, that the semantic structure of sentences and the non-referential semantic roles of their expressions are invariant across all empirically equivalent truth theories. That is, that all theories of truth for a language will take the same finitely axiomatised form, although the actual reference-specifying axioms they dedicate to each semantic primitive will depend on different though equally systematic ways of matching object language names and predicates to metalinguistic names and predicates. These variations leave the empirical content of the theories unchanged, and so this time meaning follows the evidence and we say that the meaning is unchanged. However to have meaning depend on structure for the same sort of reason would require the assumption that the same structured meanings are involved in each of the empirically equivalent theories. And this depends on the unargued for claim that there is just one structurally sound way to resolve the language into its semantic primitives. If this is an empirical claim it would have to be substantiated. If it is an a priori claim we need an argument. At this point Davidson resorts to a claim that questions of logical form will have been settled just in case we have assured ourselves that the T-theory delivers all the right T-sentences and thus enjoys an overall fit with its evidence. Certainly, it will have been settled to some extent if it is a theoretical decision taken in advance that those forms will be imposed on the evidence but this won't help us if we are trying to find out why we should see this theoretical notion as an ingredient of empirical meaning. What Davidson actually claims is somewhat stronger:

It makes no sense, on this approach to complain that a theory comes up with the right truth conditions time after time, but has the logical form (or deep structure) wrong. (Davidson 1984, p223)

But the very acceptability of this thought depends on just how one is to characterise the notion of a sentence's truth conditions. If in order to see T-sentences as giving the truth conditions of object language sentences we must suppose that their canonical proofs show how the truth value of an object language sentence depends (recursively) on its structure, then it will be no surprise that the logical forms are correct if the truth conditions are. But we are trying to motivate the notion of a semantically structured truth condition, and this just seems to presuppose it. The trouble is that truth conditions settle logical form, for Davidson, only because he uses logical form (partly) to settle truth conditions; but this is a theoretical decision and not a guarantee of theoretical correctness. We need to know why are we justified in reading structure into the evidence for meaning but not all the rest of the theoretical apparatus?

I think we are now in a position to answer this for Davidson; and the answer will rest upon a transcendental argument. Here is how it would go. The best suggestion for why it is crucial to include structure in the account of meaning is because there is just too little in the evidential points of contact with the theory to qualify as a notion of meaning. This is why meaning must be in part a theoretical notion: a matter of the imposition of semantic structure on the available evidence. It is by imposing the form of the theory on the evidence that we achieve interpretations of sentences from the very thin evidence available. It is the nature of this fit between theory and evidence that constitutes the facts of interpretation. Without the notion of semantic structure there would be no subject matter for a theory of meaning at all. So although logical form is a notion that belongs higher up in the theory, and is strictly speaking part of the machinery which helps the interpreter to make transitions between the uninterpreted and the interpreted evidence, it is to be given application to the evidence as a condition for making linguistic interpretation possible. Thus although it has no empirical support, structure *must* be imposed on the evidence; and it *can* be imposed when we go holistic and induce it over the patterns of evidence for the T-sentences by constructing a finitely axiomatised theory that enables us to derive all these T-sentences as consequences. We must do this to provide

Chapter Five

any sort of analysis of meaning at all. It is for this reason that the derivations of the biconditionals have to be taken as revealing (imposing) the form in (on) the sentences they contain. The transcendental argument, as I have called it, is an argument for semantic structure in a language, in particular, for structured meanings, and it is premised on the possibility of linguistic meaning (interpretation). Only by imposing form on the evidence which is brought to bear on the theory, can it yield rich, though empirically supportable, results of a non-arbitrary kind about the theoretical meanings assigned to sentences. And only thus can there be a subject matter for a theory of meaning.

This strategy provides us with the reconciliation we sought earlier. Only by application of an interpretation theory to a person can we find linguistic meaning in his hitherto uninterpreted behaviour; and since the evidence for meaning bears on the theory as a whole, the theoretical notion of meaning must be holistic. And by this sleight of hand, the earlier air of paradox for the truth conditional conception of meaning is dispelled. Considered one at a time, the T-sentences do not present the meanings of the object language sentences; they are entirely trivial. But when we view them as the consequences of a theory meeting the formal and empirical constraints, we come to see the T-sentences as providing correct interpretations of a speaker's language. From this theoretical perspective on T-sentences one can restore something more workable in a theory of meaning, something much less trivial than the form a T-sentence itself suggests. They do specify meanings but only in the context of a theory; and they only specify meanings *correctly* in the context of a satisfying interpretation of all the evidence concerning an agent's behaviour and conduct. But the cost of going theoretical is to make individual meanings theoretical constructs that depend on the whole of the theory in which they reside, *and* its evidential setting. Knowledge of these meanings requires grasp of the deductive apparatus whereby truth is determined by reference; but it also requires meta-theoretical knowledge about the goodness of fit between any such theory and the evidence for it. To know the meaning of any individual sentence is to be able to assign the sentence:

a semantic location within the patterns of sentences that comprise the language. (Davidson 1984, p225)

Since we cannot survey the whole of our language, we can only do this by locating the position of its T-sentences within a theory that systematises the language for us. When we know this, we know, in particular, how the truth conditions of the sentences on the left-hand sides are assembled from the semantic roles of their parts. This is just one theoretical component within a much larger complex structure. And so it is a consequence of this way of thinking, and the essential theoricity of semantical facts that meaning can only be given application only by semantic ascent. The theory plots the relations between object language sentences by tracing the truth-theoretic connections between their T-sentences; meanwhile, the criterion of interpretational adequacy for the whole enterprise "is that the T-sentences should....optimally fit evidence about sentences held true by native speakers" (Davidson 1984, p139). The advertised advantage being "that very thin evidence in support of each of a potential infinity of points can yield rich results, even with respect to the points." (Ibid., p137); and these "rich results" are available only to interpreters. We need to know a theory to be able to construct meanings; but we need to know much more about a theory to be able to construct interpretations. It is only by knowing a formally satisfactory theory that we can move from the primitive data to interpretations of sentences; but it is only by knowing enough about that theory and its relation to the evidence that we can know that any particular sentence's truth condition *is* an interpretation.

5.5) T-theories, Sentence Meanings, and Individual Beliefs

We can now review Davidson's Semantical Minimalism. It arises from the need to find an empirically constrained notion of meaning available to anyone with the resources of the radical interpreter. The concept of meaning cannot be purely theoretical and empirically unconstrained, or else we could make anyone's words mean anything we liked. At the same

Chapter Five

time there is no hope of constraining our ascriptions of meaning so as to keep faith with the speaker's intentions in uttering. For although a speaker's meaning must conform to his intentions, from the interpreter's point of view it conforms to the intention the interpreter attributes to him; and in attributing an intention, the interpreter is largely guided by an understanding of his subject's speech. The evidence for meaning and intentions comes from evidence that mentions neither. Faced with evidence that cannot by itself support the notion of meaning, the interpreter needs to manufacture meanings and beliefs by an interpretation theory that gives them joint application in the re-described evidence of an agent's behaviour. The theory of linguistic interpretation is a sub-part of this overall theory. Now to find an empirically respectable notion of meaning which respects the fact that meaning is theoretical and that evidence for meaning bears on interpretation as a whole, one must see meaning (interpretation) as the result of bringing the empirical evidence under the control of a formal truth theory. And although not every aspect of the theory is an aspect of meaning, there is just too little at the different points of contact between the theory and its evidence to constitute meanings, so the possibility of meaning requires that what we find at those points are sentences with a semantic structure induced by theory over a pattern of truth values evinced from a speaker's attitudes and actions towards sentences and utterances. Since the parts revealed by those structures cannot play any substantial part in interpretation, it is only the structures of these truth-evaluable sentences that constitute their meanings. These meaning-structures give us what Davidson calls a sentence's role in the language. This has to be more than its truth-value, since that alone does not tell us why it depends on the patterns of truth-values other sentences can take; but it is much less than a means of referring to the places, people, and things in the event, or state of affairs that the sentence is about:

The question what objects a particular sentence is about, like the questions what object a term refers to, or what objects a predicate is true of, has no answer. (Davidson 1984, pxix)

Chapter Five

To account for the individual meanings of sentences is to account for the role each sentence plays in the language, which means considering the semantic dependencies in the context of all other truth-evaluable sentences of the language. Since, for Davidson, it is the structure of a sentence that determines its relations to others; it is these structural relationships which give it "a semantic location in a pattern of sentences that comprise the language". This, then, is what Davidson understands sentence meaning to consist in. The analysis of meaning is the analysis of semantic structure; a structure best accounted for by a theory of truth:

To see the structure of a sentence through the eyes of theory of truth is to see it as built up by devices a finite number of which suffice for every sentence. (Davidson 1984, p55)

And so it will be a theory of truth which supplies the correct theory of meaning:

I suggest that a theory of truth for a language does, in a minimal but important respect, do what we want, that is, gives the meanings of all independently meaningful expressions on the basis of an analysis of their structure. (loc. cit.)

Since the meanings of sentences are given in terms of their structural relationships to all other sentence, the structures, and hence the meanings, we can assign to them are constrained holistically by the evidence that fixes their T-sentences. Evidence about these should reveal a pattern of truth values amongst the object language sentences. Therefore, the identifications of *sentences* and the truth-values speakers attach to them is vitally important to this whole approach since it is on these that semantic structure is induced by appeal to the apparatus of a truth theory that entails all those T-sentences. The structures induced and therefore the meanings of every sentence depend on the totality of evidence. To mis-attribute sentential attitudes to a speaker, by mis-identifying his sentences, or his assent or dissent to them, can lead one to make false assignments of meanings and therefore to mis-identify a speaker's language. Davidson has given us no cause to think that sentential attitudes are any more

reliably attributed or immune to revision than any others, and yet it is on these that the possibility of a more precise specification of an agent's mental life depends. What would help to stabilize this part of interpretation theory is a clear identification of the individual sentences involved. But it is difficult to see what the syntactic identification of one among many meaningful sentences consists in for the interpreter since these identifications are brought to light by the re-descriptions of the "mere emissions of noise", and their identifications depend on the cogency of the overall pattern revealed by interpretation in which they are located. More precise syntactic criteria of identification are nowhere available in the interpreter's picture. If it is assumed that we need a recursively specified syntax for the declarative part of a language just to identify the interpretable and truth-evaluable sentences (whether or not we recognise semantic structure in them in the purely homophonic case), then it is going to be hard to motivate a speaker's, or interpreter's, knowledge of this, even by rational re-construction. For while there is a perfectly coherent method for extracting the semantic features of words from sentences by rational inductive means, this depends on a suitable presentation of the linguistic data of sentences and dependencies, which in turn rests upon certain syntactic information, and here it is just not obvious what the rational inductive means of describing syntax could produce. What are the premisses; what inductive inferences we should draw? Grammarians have long argued that the evidence available to the ordinary language learner is just too impoverished to support the generalisations necessary for the notion of syntactic well-formedness, and yet all speakers have it. Famously, Chomsky draws the conclusion from the "poverty of stimulus" arguments, that high-level grammatical generalisations must be innate. Without this move there is a vast gap to be plugged in the account that tells us how we get from sounds to meanings; a gap that must be filled by an account of how we construct a descriptively adequate grammar for our language that meets all available empirical criteria. Candidates for this role are noticeably absent. The proper conclusion of all this should be that whatever little there is to the notion of meaning on the Semantic Minimalist's picture, his claim to

Chapter Five

recover it from the primitive behavioural data by means of an interpretation theory alone looks highly implausible.

Returning to the minimalist's conception, it is also worth noting how the holism it enshrines guarantees a highly theoretical status to the notion of individual sentence meanings, which in turn makes knowledge of meaning too theoretical to capture the facts about ordinary speaker's linguistic understanding. It seems as if this vital part of any person's mental life is either unjustifiably ignored or misrepresented. The point emerges as follow. It is the totality of T-sentences derivable within the theory that fixes the interpretation of any particular sentence dealt with by one of them. So it is not just the T-sentences and their canonical proofs that give a meaning (interpretation) for a sentence. It is the whole theory with all its logical consequences. The fundamental conception of meaning is one of a theoretically suspended item tied to the evidence via the rest of a theory. If the meanings of individual sentences are specified by reference to the whole of the theory, we shall need the whole of the theory to specify their meanings each time. This is just what Davidson expects:

we can give the meaning of any sentence (or word) only by giving the meaning of every sentence (and word) in the language.
(Davidson 1984, p22)

There are an infinite number of sentences to take into consideration, and finite creatures like us can encompass them all only by means of our knowing a theory that recursively characterises each of them. In this way, the distinctive role in the language of any particular sentence will be acknowledged when specifying its meaning if this is done by means of a theory that already anticipates, though its deductive apparatus, the relation of that sentence to all others with which it deals.

There may be a number of different theories one could construct compatibly with the evidence. Building the theory, in a piecemeal fashion means that "a pattern must be built up that preserves the formal constraints while suiting the evidence as well as may be" (Davidson 1984, p152). Interpreters may satisfy this requirement while arriving at different interpretational schemes; but their use of different theories need

Chapter Five

not threaten the idea that they can know the meanings in a particular public language. This idea can be preserved by the semantical minimalist:

I suggest that what is invariant as between different acceptable theories of truth is meaning. (Davidson 1984, p225)

This will be a field of semantical structures, each one of which is a theoretical component knowable only on the basis of knowing a theory to which it belongs. It is by reference to this notion that Davidson fixes the identity of languages, and not by the schemes of reference embodied in particular theories of interpretation. So as far as the character of the language goes this is indeed significant semantic structure (subject to the misgivings if we are not guaranteed a unique assignment function pairing these structures with individual sentences.)

Davidson is now entitled to say that meaning reflects the structure of the language, and to that extend, the structure of language use. But remember that the structure reflected in meaning is the structure of the theory, and there is a lingering temptation which, Davidson would have us resist, to ask whether the theory is *correct*. As we know Davidson would simply deny that there is anywhere else to look for an answer save in the formal and the empirical constraints on truth theory. The formal constraints are justified by the requirements of interpretation, and the empirical constraints just provide us with too little to say without appeal to the theory.

Now of course the confirmation of T-sentences does depend on belief-ascriptions, as the other contributory factor to the evidence of speakers holding sentences true, and so this means that truth conditional assignments are not so hopelessly unconstrained as just to arrive at the correct overall pattern of sentences' truth-values. But remember these belief-ascriptions are made from within interpretation theory and such ascriptions themselves depends for their precise rendering on the way we can interpret the believer's language; so as far as matters of semantic refinement go they may not constrain logical form and semantic content that much. Certainly, we have to interpret people so as to make the sentences they use relate to one another in much the same way that their

Chapter Five

thoughts relate to one another, and all this in a way that conforms to - ie. depends on and constrains - the speaker's behaviour and attitudes in these and other linguistic situations; but this tells us little about how fine-grained or course cut the contents of someone's utterances and beliefs should be, or how much indeterminacy is tolerated in a *correct* interpretation of him. All we have from Davidson is a policy decision to adopt a Tarski-style theory of truth for interpreting a language, which enjoins a finite axiomatisation and a wide range of logical forms but does not guarantee unique assignments of forms to individual sentences of the language. And if matters of logical form have not been settled just by taking that decision alone, then there is room to cast doubt on Davidson's assumption that logical form is (nearly enough) invariant across different acceptable T-sentences. But the indeterminacy point aside, Davidson's attributionist picture simply produces a point at which justification gives out and where we are dissuaded from asking any more. At that point the best Davidson can do is try to persuade the recalcitrant that:

if we were to ask for evidence that the explanation is correct, this evidence would in the end consist of more data concerning the sort of event being explained, namely further behaviour which is explained by the postulated beliefs and desires. Adverting to beliefs and desires to explain action is therefore a way of fitting an action into a pattern of behaviour made coherent by the theory. (Davidson 1984, p159)

The appeal to 'coherency' like the appeal to rationality represents an ultimate stopping place for inquiry. The only further thing to be said in favour of the theory is to speak of what Frege would have called its 'fruitfulness':

if we ask how a method of interpretation is tested. In the end, the answer must be that it helps bring order into our understanding of behaviour. (Davidson 1984, p161)

But it is at just this point that one can wonder whether it does. Do we ordinarily use such a theory to make sense of others? Further, we can

wonder whether it can tell us enough about what we need to understand when conversing with others: in particular, what they are talking about. These are both objections to Davidson's picture based on an appeal to the notion of speaker's knowledge of language. If this notion can be substantiated it will in no small respect threaten the whole foundation of Davidson's attributionism. For what seems so unsatisfying about that picture is that it leaves meaning not so much in the mind of the speaker, but that it leaves both in the eye of the interpreter who interprets what is said. Meaning is the product of applying a formally adequate truth theory to the behavioural evidence of what a speaker says and does:

Such a theory may be taken as giving an interpretation of each sentence a speaker might utter. To belong to a speech community - to be an interpreter of the speech of others - one needs, in effect, to know such a theory, and to know that it is a theory of the right kind. (Davidson 1984, p161)

In the terms used above, this might seem an unnecessary conflation between the interpreter and the theory of what the interpreter knows. I shall turn to this next. But note that on Davidson's conception of interpreted meanings it is quite unclear how the speaker is to *produce* his linguistic responses. Perhaps the proper conclusion here should be that there is no guarantee that the ordinary speaker has an understanding of the structure of his own language. Being a speaker would not automatically give one insight into the character of one's own language since this would include structure constituted by a theoretical characterisation of the language. However, taken to extremes this would mean that the speaker does not have an understanding of his own language: which is absurd. To avoid this, Davidson has recently written:

an interpreter uses his theory to understand the speaker; the speaker uses the same (or an equivalent) theory to guide his speech. For the speaker, it is a theory about how the interpreter will interpret him. (Davidson 1986, p438)

This is an extraordinary claim. We may well have a theoretically developed view about how we will be understood, but it is a huge jump from there to the claim that this theory *guides* us in saying what we do. The psycholinguist has an extremely complex view about the structure of the sentences he utters and the reception properties of his hearers, but he could not claim in any uncontroversial sense that he was using that theory to *guide* him in producing speech. The controversial sense would talk about the psychological reality of the theory and suggest that rather than the speaker using the theory, the theory is true of those unconscious processes in the speaker causally responsible for his speech. The corresponding claim for the psychological reality of truth theory would argue that the rules and representations that make up the deductive apparatus of a correct truth-theory should describe (at least some part of) the actual psychological mechanisms which are causally responsible for the linguistic behaviour of competent speakers. But as we saw above Davidson eschews the explanatory pretensions of cognitive psychology. So we are still owed some account of what it is to be guided by such a theory. It is here that Davidson needs to recognise Wittgenstein's *Auffassung*: "a way of grasping a rule that is not a matter of interpretation".

The trouble with the attributionist picture is not that it has nothing to say about the psychological reality of the interpreter's semantical pronouncements for the speaker: in its re-working of that notion, the psychological characteristics of the mind do confirm the ascription of meaning and structure, but it is just that they too are externally attributed from the third-person standpoint of the interpreter, which means our subject's psychological reality is partitioned and exhausted by the conditions on attitude-ascription, conditions which include interpretations of his speech. The trouble is that there is no independent check on the psychological reality of an externally attributed semantic theory. But then, there is no such thing for the theory to conform to, as far as the attributionist is concerned. Neither has he any other independent means of confirming his theory: there are no "physical underpinnings for the deductive apparatus of truth-theories, over and above their output" (McDowell 1978 p126), and for the attributionist this exhausts the places to

look: there is only the explanatory space of physical theory that provides causes of the way the world works, and the explanatory space of intentional theory that explains behaviour, understood in terms of the sort of thing an agent could see reasons for going in for. (See, McDowell, p126). Now since interpretation theory does fit into the patterns of causal explanation in physical theory, Davidson's talk of a speaker being guided by such a theory must belong to the latter type of explanations. But this is wildly disruptive of the normal picture of unreflective language use which McDowell advocates so strongly. I shall return to this in the next section.

It has been a constant criticism of this approach, by antirealists like Michael Dummett and Crispin Wright, that it gives us no account of what a speaker's knowledge of truth conditions consists in. But now we are better placed to see why Davidson omits to give this. From the attributionist standpoint there is nothing further to say over and above providing an account of what it is for the sentences in the speaker's language to have particular truth conditions assigned to them. Confirmation that the theory gives the meaning of the sentences in question, and that they are publically knowable, is offered in terms of the theory's success: that knowledge of the theory would suffice for understanding the sentences of that speaker's language. Any further confirmation is either impossible or resides not in the nature of speaker's knowledge itself, but in the theory which entitles us to so describe him in terms of his beliefs, desires and reasons for action. The structure and content of the rational mind, like the structure and content of the language it so crucially depends on, are externally attributed by the theory of interpretation. What then becomes so crucial in evaluating these theories is their precise nature of fit with the behavioural and physical evidence. But as we now know, it is only via the a priori constraints that the physical conditions become an important part of the grounds for the attributions made to a subject in interpretation theory. Unharnessed ascriptions of truth conditions to a speaker's sentences that make him speak about remote parts of the world, or in principle unknowable

Chapter Five

conditions, are meant to be reined in here. But this all depends, once more, on the resources of the interpreter.

But still one wants to protest that not every external attributed theory correctly captures the psychological and semantical facts about an individual mind and an individual's grasp of his language. Saying more precisely which ones do is vital whenever we are faced with philosophical controversy or scepticism about the nature of the attributions as in the realist-antirealist debate. Davidson's stopping place just won't do. Even the transcendental argument I ascribe to him will not settle the more controversial issues. What we would need to know is whether the conclusion of the transcendental argument tells us something about the features of reality, in this case, the reality of linguistic structure, or merely tells us what someone with that conceptual framework must think. Conceived in the latter way, it is not surprising that it must make the structure-positing move to retain any credible notion of meaning. But to the non-attributionist, this will look like a desperate attempt to come to grips with explicable facts about actual languages and actual language users. We only need look in the right place.

The problem that faces us is to motivate an account of semantic structure and the semantic properties of sentence-constituents which is less blankly external to the minds and linguistic practices of the ordinary speaker. On the attributionist's picture these properties of individual sentences have no psychological reality or empirical significance over and above that which attaches to the theory as a whole. The holism of interpretation theory denies these key notions any status beyond that of abstractions from the overall intentional organisation of a speaker's behavioural history. What a non-attributionist psychology must make room for is some more local correlation of these semantic features of sentences with properties of speakers' knowledge or mental states. This will require an adequate conception of both the object of knowledge and the conditions for possessing it. Clearly, Davidson's notion of a sentence's meaning or truth conditions won't serve as objects of knowledge of this kind since they are theoretical components belonging to a complex

Chapter Five

structure. We need an alternative conception of sentence meaning, and the best place to start is with a Fregean conception of truth conditions.

For Frege, the truth-values of sentences were determined by the references of their parts. But speakers working out the truth-value of a sentence from the references of its parts would have to go via the senses of those parts, for speakers never grasp just the reference of an expression (and in the case of intelligible but undecided identity statements they do not grasp as much as the reference of an expression) except by way of the sense of the expression of which that thing is the reference. Also for Frege, sense is a cognitive notion, a notion we use to explain what it is a speaker understands by an expression, or the meanings he attaches to it. So the determination of a sentence's truth-value on the basis of its parts is a possible object of speaker's knowledge - to know that sentence's meaning is to grasp its truth conditions. Thus truth conditions belong at the level of sense and can be regarded as parts of a speaker's linguistic knowledge. Since the sense of the sentence, that is, its truth conditions, involve a number of subsentential senses the sense of the sentence must be structured, having those senses that determine the relevant truth-affecting references of words as its constituents. Thus semantic structure is also part of speaker's knowledge. Thus Frege's conception of truth conditions gives rise to a very rich conception of speaker's knowledge. As Dummett describes it, Frege's requirement on understanding builds into it the notions which Davidson locates higher up in the structure of the truth theory:

For Frege, a speaker grasps the sense of a sentence only by apprehending its truth-value as being determined in a certain way corresponding to its structure, the senses of the constituents determining their referents and hence the truth-value of the whole. The conception of the references of the parts of the sentence therefore enters into the speaker's grasp of the thought it expresses: without this conception, we cannot explain what it is for him to know the condition that must obtain for it to be true. (Dummett 1981, p461)

This strong requirement would have it that to understand an indicative sentence, one must have a conception of the conditions for the sentence's truth that essentially involves grasp of its structure and knowledge of what it is for certain things to be the referents of its terms. This means that one is constrained to find a theory which accurately portrays the referential content and structure of a speaker's knowledge. To come up with the correct Fregean truth conditions of a sentence is most certainly to fix the correct logical form. But to do this one must find the right semantic structure for a speaker's sentences and the right reference for their structurally revealed parts; and these constraints of reference and structure on the correct theory of meaning for a speaker's language are constraints on correctly describing what that speaker understands by the sentences of his language. His understanding of his own sentences is a matter of what he knows, for the meanings of his sentences are genuine objects of knowledge. In this sense an adequate theory of meaning must be a theory of understanding; a theory of sense must be a theory of a thinker's grasp of sense.

This Fregean conception of truth conditions is not applicable on Davidson's picture, and this is where McDowell's defense of Davidson looks strained. The notions that Frege sees as integral to a sentence's truth conditions as genuine objects of speaker's knowledge, Davidson sees as merely posited to implement the theory. Even in attribution there is no suggestion that these notions of the semantic machinery must enter the mind of the speaker. Unable to make use of these constraints on an adequate account of meaning for a language, Davidson has no way to cut down on the threatened indeterminacies. But he shows his semantic minimalism in full by embracing the limitations of his linguistic theory:

This doctrine of indeterminacy of translation, as Quine called it, should be viewed as neither mysterious nor threatening. It is no more mysterious than the facts that temperature can be measured in Centigrade or Fahrenheit (or any linear transformation of these numbers). And it is not threatening because the very procedure that demonstrates the degree of indeterminacy at the same time demonstrates that *what is determinate is all we need*. In my view, erasing the line between

analytic and synthetic saved philosophy of language as a serious subject by showing that it could be pursued without what there cannot be: determinate meanings. (Davidson 1986, p313 emphasis mine)

Now combine this view with the thesis of the inextricability of belief and meaning: "the only access to the fine structure and individuation of belief is through the sentences speakers and interpreters of speakers use to express and describe beliefs." (Ibid., p315) and we have a position that we could now think of Davidson's position as a combined **Semantical and Psychological Minimalism**:

What a fully informed interpreter could learn about what a speaker means is all there is to learn; the same goes for what a speaker believes. (Davidson 1986, p315)

Total theories are what we must construct, and many theories will do equally well. (Davidson 1984, p241)

Indeterminacy of meaning or translation does not represent a failure to capture significant distinctions; it marks the fact that certain distinctions are not significant. (Davidson 1984, p154)

This is a high price to pay if it means that the distinction between thinking about this person rather than that one is not significant, or cannot be made for the subject at all. But this is just what the interdependence of language and the propositional attitudes leads to when combined with the inscrutability of reference. The traditional notion of the intentionality of a mental state, the notion that there is something a particular thought is about is dispensed with. A further consequence is that if the structure and content of language provides the only guide to the structure and content of thought, then Davidson will have to say that the linguistic articulation provides some guidance to the conceptual structure of thought, but that the interpretation of language undermines the idea that the distinct conceptual constituents contribute to determining the contents of the thoughts in which they partake. The conceptual structure of a thought - subject to whatever structural indeterminacies there are in

language - offers no guide to parts which determine the referential features of a thought, for there is really no such thing. What else, then, could the conceptual constituents of thoughts be *about* ? For Davidson there is little else we can say about them. So the concepts we might reasonably suppose make for the precise contents of our thoughts will have a purely instrumental status in the context of belief attribution. There is just no fact of the matter concerning questions about whether or not someone possesses this or that particular concept, save whether he is apt to be ascribed it in an overall interpretation that makes best sense of his behaviour. Davidson says that the "question of concept possession is inseparable from that of attitude ascription" (A&E, p221), and in a perfectly respectable sense this is true. But this picture encourages us to go further in thinking that there is no empirical significance to concept-ascription. This should be denied.

To sum up, the principles of rational psychology and semantic theory are constitutive of the practice of interpreting others' behaviour, and since there is no gap between the best interpretation we can produce and the mental life of the interpreted subject, these principles are constitutive of the concepts of mind and meaning. The mental and the meaningful are introduced into the world by way of these concepts because they are the notions we use to give an intentional organisation to things that are really physical. Minds and meanings are manufactured from the available evidence brought under the control of the principles of interpretation, and permit the degree of discrimination or indeterminacy that attaches to the language we use to report another's speech or thought. Save for the principles of interpretation, there is no other way to cut down on the indeterminacy of language or thought. The beliefs an interpreter can ascribe to someone will be identified by their contents, and the contents of beliefs will be inseparable from the constructions placed upon sentences held true because of them in theories of linguistic interpretation. However, the assignments of meaning made here cannot be separated from attributions of belief; beliefs and meanings form an interconnected system, much like the theory that ascribed them, which confronts the evidence as a whole. Every attribution of an attitude, or interpretation of a

sentence, is a move within a holistic theory for interpreting a subject's behavioural output. This means that language, like the sentence meanings it contains, and the individual beliefs identified by them, is a largely theoretical construct imposed on the available evidence: the characteristic properties of a language are identified by the interpreter and attributed to the words of a speaker. This is achieved by the theory of truth. Given an antecedently understood notion of truth, and a theory's fulfilment of Convention T, the constraints which do most of the work in ensuring that we have an interpretative truth theory are the **Propositional Attitude Constraint** and **Systematicity Constraint**. The first brings with it a holistic constraint on theories of interpretation, and the second makes sure the embedded theory of linguistic interpretation can take advantage of a compositional molecular semantics; at least in a theory-internal way. If theories of meaning meeting both of these requirements leave room for a great deal of indeterminacy, then we must either look for further adequacy conditions, or embrace Semantic Minimalism. The indeterminacies we discovered are that equally adequate truth theories may assign different truth conditions to the same sentences if they make compensatory adjustments in the speaker's beliefs. This is possible because beliefs and meanings are dependent variables in a theory of behaviour built on irreducible, intentional notions that enjoy a holistic fit with the physical evidence, and it is a feature of this total fit that various instantiations of the dependent variables can be realised by the same physical base. So theories making different attributions can be applied to the same individual. Secondly, the contents of his utterances and beliefs are referentially inscrutable. And thirdly, different truth theories may assign their semantic structures differently to the same set of sentences.

Alternatives to this attributionist conception of mind and language may be motivated by the thought that one must cut down on this degree of indeterminacy and restore facts of the matter about what people think and say, including facts about whom and what they are speaking and thinking of. This sort of opposition to attributionism would be motivated by a methodological maxim to the effect that if there is a way to reduce the indeterminacy in our best account of meaning and mind without lapsing

into philosophical error we should take it. This would be motivation enough, but I think we can make a stronger claim about the inadequacy of the attributionist's picture. For even if the meanings attached to individual sentences cannot be constrained by referential properties of their parts, they should be constrained by the demand that a theory of meaning should provide an account of speaker's knowledge of meaning. And in this respect, Davidson's account is deeply flawed; for he can be seen to be ascribing to speakers/ interpreters less than they know, while on the other hand crediting them with knowledge they do not have. This suggests that he is failing to recognise important facts about speakers' knowledge which a non-attributionist account of these matters would have to make out. Moreover, even in the case where we accept the minimalist conclusions about meaning and accredit the interpreter with reflective knowledge of the theory, it is doubtful whether he can recover the propositional forms used in truth theory from "mere emissions of noise" without resort to a descriptively adequate syntax. The attributionist position on knowledge of syntax will look much less plausible once we see what sort of theory is required.

I now want to motivate a more explanatory approach to matters of mind and meaning, based on a different conception of psychology which will lead to the construction of more substantial psychological constraints on theories of meaning. I shall suggest that in the end the correct way to substantiate the psychological reality of individual sentence meanings is by reference to cognitive psychological facts about language users. However, to be a non-attributionist all one is required to believe is that there are facts of the matter about the psychological lives of speakers and the linguistic properties of their languages which are there to be *described*, rather than *constituted*, by correct theories of them. Those who believe this while eschewing all empirical psychological investigation, will have to establish the existence of a domain of facts about a speaker's knowledge of language which are not just a matter of attribution and which, being confirmed independently of their ascription to a speaker, provide some constraints on what that the theory can legitimately describe a speaker as knowing. The less cognitive psychologically committed picture of

speaker's knowledge will still have to face up to the challenge of describing a speaker's knowledge of syntax and say what it is for him to possess that knowledge. I shall claim that the most plausible approach to matters of syntax is to be given by an explanatory theory that describes facts about a speaker's psychological organisation that explain aspects of his linguistic competence.

Both of these approaches to a speaker's knowledge of language will try to provide substantial constraints on the theoretical assignments of semantic structures and semantic properties to sentences. And whether the desire is to describe aspects of the psychology of the language user, or simply to provide further constraints on a semantic theory adequate for describing the determinable properties of an actual language, the basis for claims about properties of structure and reference in a language must go via speakers' knowledge of those properties if we are to reject the model-theoretic approach to semantics I rejected at the outset. The flaw in the Davidson programme is that it lacks the resources to capture these essential linguistic facts and misdescribes the character of linguistic understanding.

I shall now turn to these alternative conceptions, and I begin by contrasting the resources of the explanatory, the descriptive and the interpretive theories of meaning.

Chapter Six

Philosophy, Mind and Language

6.1) Theories of Meaning, and Theories of Understanding

In a number of papers, Michael Dummett has insisted that questions about meaning are best interpreted as questions about understanding; and since he also believes that questions about meaning should be tackled by attempting to construct a theory of meaning for a language, he has advocated that we should think of a theory of meaning as a theory of understanding. Let us call this Dummett's equivalence claim. It says that a theory of meaning should be an account of what a speaker knows when he knows how to speak a language. Dummett tends to stress that the account of linguistic knowledge should be 'speaker-centered':

What we are after is an account of the sort of understanding a speaker has. (Dummett 1976, p69)

This account of what it is to know a language will be a detailed account of what it is to have the knowledge a speaker has about his own language; in particular, it must state what he knows about any given sentence.

Now we must take care when interpreting Dummett's dictum for it can be read in different ways. At its most bland, it seems a truism that the meaning of a sentence is what someone who understands that sentence knows, and unless one is prepared to say more, it seems to add little to describe a theory of meaning as equivalent to a theory of understanding. But Dummett does add more, not just by stressing that the theory of meaning should account for the nature of speakers' knowledge, but also by insisting that it should also tell us what a speaker's having this knowledge consists in. That is, it should be possible to say what someone who possesses that knowledge has that distinguishes him from someone who lacks it. This extra requirement turns the equivalence into a knowledge constraint on theories of meaning. The force of this constraint will depend on what one is prepared to take as establishing that a person possesses any given piece of linguistic knowledge. As is familiar, Dummett is prepared to ascribe someone a knowledge of the meaning of a particular sentence (a grasp of its truth conditions) only if that person can distinctively manifest

Chapter Six

this in behaviour describable in terms other than his saying that such and such is the case. Indeed the behaviour in which someone manifests his understanding of a sentence should be recognisable by another without command of the language in question; for only then could another learn what the sentence means (on that speaker's lips) without merely guessing this, or knowing it already. The sort of thing envisaged as manifesting this knowledge is a complex discriminatory skill exercised in response to publically accessible circumstances. (see Wright's "Strawson on Anti-realism" in Wright 1987) This version of the further requirement on knowledge - that speakers be able to show what they know - is the manifestation requirement, and it is this which turns Dummett's knowledge constraint on theories of meaning into an antirealist requirement. For there is nothing in a speaker's ability to use a sentence with purportedly verification-transcendent truth conditions that shows his sensitivity to the obtaining or not of those conditions; so there is nothing to show that he has that knowledge. Indeed someone's grasp of the difference between the obtaining or not of those truth conditions could not be shown to an audience with a merely human range of cognitive capacities, for it would transcend their checking powers too, *ex hypothesi*. So instead of treating the competent speaker as manifesting knowledge of this sort, Dummett regards what he is capable of manifesting as being his sensitivity to a whole range of evidential conditions which bear upon the warrant for the assertion or denial of that sentence. This would amount to displaying knowledge of the sentence's assertion conditions.

I have said that the equivalence Dummett sees between a theory of meaning and a theory of understanding is just the claim that a theory of meaning must be a theory of speakers' knowledge of meaning, and that it is a further demand that the theorist should be able to say what it is to *possess* the items of knowledge the theory describes. But now I want to look at different ways of reading Dummett's equivalence claim. In fact, there are at least three ways to read the claim that a theory of meaning is a theory of understanding. We can read it from left to right, from right to left, or without giving any explanatory priority to either side.

The attributionist about linguistic knowledge, like the Platonist about meanings or senses, will read it from left to right. That is, he will consider

Chapter Six

it enough to give an adequate account of what the sentences in some speaker's language mean, without looking for further evidence that these are the meanings the speaker of the language actually grasps. This theorist will not equate a specification of what a speaker knows with an attempt to account for a speaker's mastery of his language. If the theory of meaning adequately characterises the language in question then these will be the meanings speakers of that language grasp: ie. the claim for correct linguistic specifications, here, will licence the ascription to speakers of the corresponding linguistic knowledge. This sort of theorist feels under no obligation to say more about a speaker's having that knowledge than he says by describing the objects of that knowledge. In particular, he will not feel the need to say what it is for the speaker to acquire or exhibit an understanding of his language. But the equivalence is maintained since the theoretical description of what sentences in a speaker's language mean will suffice to give a partial characterisation of his mind. Here, as in all cases of attributionism, we are invited to think that there is no further fact for a speaker's possession of that knowledge to consist in.

Another way to read the requirement is to see no priority between the right and the left hand side. To describe one side is to describe the other. If we say what someone knows when he understands an expression or sentence, then we will be saying what that expression or sentence means, and vice versa. But for this theorist, unlike the first, there will be no way to arrive at an adequate theory of meaning without including satisfactory conditions on what it is to know the objects described. Meaning is an object of knowledge, so a theory of meaning must be an account of items of knowledge. This makes epistemological matters crucial to semantic theory. At the end of the day, someone of this persuasion will think that the two sorts of theory are really just one: a theory of meaning just is a theory of the knowledge a speaker has.

Although this theorist sees meaning as an *epistemologically* constrained notion, he will not see an account of meaning as answering to *psychological* constraints. This is what distinguishes him from the third of our protagonists. The no-priority theorist (like the first protagonist) can see no epistemological asymmetry between the task performed by a theory of understanding and that performed by a theory of meaning. There is no

extra role for a theory of understanding to play in constraining choice of an adequate theory of meaning, and so no extra purchase it affords us. The only candidates for being theories of meaning for a language must already be theories of understanding, and the only form a theory of understanding can take is the form given to it by providing a theory of meaning that delivers meaning-specifying theorems of the familiar kind.

The right to left reading of the equivalence would give explanatory priority to the right hand side giving the theory of understanding a more substantial role to play than either of the above two positions. This would not entail a reduction of meaning to an account of a speaker's non-linguistic intentions but it would stress the relevance of certain psychological considerations to any account of linguistic meaning. For on this position, the theory of understanding would be embedded within a wider context of the psychological states and processes, and overall cognitive organisation of the language user. So for example, a theory of meaning might be required to contribute to the psychological explanation of a speaker's linguistic behaviour. Our best model of psychological explanation would then impose strong constraints on the left-hand side of the equivalence claim. For instance, where psychological explanation is a form of causal explanation, and where the causal processes are computations, the axiomatic structure of a theory of meaning might be required to reflect those states of a speaker that are implicated in computations that result in the various observable linguistic outputs. On this approach, the *form* of a theory of meaning would be a matter of empirical hypothesis. The information states of a speaker, his information-processing and perceptual capacities, his relations with the environment and with others could all contribute to determining the meanings he could grasp. This would amount to giving a substantive account of understanding; one which says not only what a speaker knows but also how he acquired that knowledge, how it is encoded, and how it is deployed. The epistemological asymmetry would be due to the fact that empirical theories would aim to reveal antecedent conditions on learning and using a language which could then shape the notion of meaning the theorist of meaning could legitimately use to describe the language of his subject.

Chapter Six

As well as drawing upon the psychological conditions for understanding, such an account could also draw upon (without reducing meaning to) the notions of content in the theory of thought. But Dummett's equivalence operates at the level of thought too: a theory of thought must be a theory of our grasp of thought. This too is open to various interpretations. Someone could suppose that a logical account of thoughts and the relations between them would be an account of what thinkers could grasp. Others would insist that we need to give some substance to the epistemological claims being made about just what speakers grasp. But someone who thinks there is a point in looking for some non-attributionist theory of grasp need not conceive this as a psychological theory. The advocate of this middle position would try to distinguish between giving a theory of our grasp of thoughts and giving a theory of thinking. He would see the former as the proper task of the philosopher and the latter as the proper task for the psychologist. Like the Platonist, someone holding this position will want to distinguish his logical inquiries from a psychological inquiry; but unlike the Platonist, he will believe that logical claims involve epistemological matters about the precise nature of the thoughts speakers actually grasp. The epistemological constraints on the logical account of thoughts will not, he thinks, involve any psychological constraints. Distancing the theory of grasp from psychological theorising depends on making use of some such distinction as that "between the process by which we come to acquire a grasp of sense and what constitutes such a grasp" (Dummett 1973, p240), where only the latter is of interest to the philosopher of language. But this depends on coming up with an account of what constitutes grasping certain thoughts; an account which does not just read-off facts about grasp from a plausible and coherent theory of thought, but which is also an account that eschews all psychological facts about a speaker detectable by empirical research. However, our third protagonist will regard success in the theory of thought as very much dependent upon psychological facts about our thought-processes, such as whether or not they involve the processing of structured states. To this theorist, any facts of the matter as to whether or not we grasp something structured will be referred to independently

Chapter Six

investigable features of our psychological processes, such as whether their inputs, outputs and intermediate working structures have a syntax.

The notion of grasping a thought is Frege's, as is the analytical conception of a study of meaning or thought. Although Frege never fully dealt with the notion of grasp in a satisfactory way, his remarks in the following passage are illuminating. Speaking of our ability to grasp a thought about a law of gravitation he wrote:

... it is a process which takes place on the very confines of the mental and which for that reason cannot be completely understood from a purely psychological standpoint. For in grasping the law something comes into view whose nature is no longer mental in the proper sense, namely, the thought; and this process is perhaps the most mysterious of all. But just because it is mental in character we do not need to concern ourselves with it in logic. (Frege 1979, p145)

The Platonist sentiments expressed in the final sentence risk denying Dummett's equivalence between thought and our grasp of thought. But it is interesting to note that attempts to re-introduce it in the context of Frege's view would involve us in an inquiry on the borders of philosophy and psychology. This is just what the explanatory theorist believes.

Notice that a Fregean Platonism which attempted to eschew all epistemological matters about speakers' grasp would be quite unworkable as a theory of sense (thought), since the individuation of the senses of words or sentences, and the concepts exercised in judgements depends in part upon their cognitive significance for creatures like us. A purely logical theory of sense that prescind from all facts about the psychological states of thinkers still has to recognise a notion of cognitive value that connects the finely individuated senses of thought and language with facts about people's ordinary propositional attitudes. Judgements of identity or distinctness among the conceptual contents of thoughts or sentences are sensitive to facts about whether a thinker can take different attitudes towards those contents of sentences. No acknowledgement is made by the Platonist of why the precise degree of difference between the logical possibilities of mind-independent senses should coincide with the degree of discrimination in psychological attitudes reflected by human thinkers

Chapter Six

in reports of their own or others' mental lives. The third kind of theorist would simply insist that the only thoughts there could be would be those which fell within the human range of cognitively discriminable judgements. He would insist that thoughts simply are what a theory of our grasp is a theory of.

Let us now review these three positions. We can think of someone who occupies the third position as a cognitivist, committed to giving an explanatory account of our intellectual achievements, such as our capacity for inference, perception, language use and understanding. The cognitivist need not believe that all the facts about the contents of mind or language should be given in terms of the psychological apparatus of the language user (thinker); nor need he believe that a theory of meaning (content) is no more than a sub-branch of a scientific individual psychology. He can claim that while we need to know something of the internal cognitive organisation of language users, any adequate account of intentional contents would have to mention relations between states of a thinker's cognitive system and objects in his environment, or draw upon the social aspects of communal language use. None of this is ruled out by aversion to the psychological underpinnings. But what a cognitive theorist of language does insist upon is that the intentional level account should incorporate the structural details of the cognitive system that makes possible someone's grasp of the relevant contents. The citing of certain crucial features of a thinker or speaker's cognitive psychological organisation is a necessary condition on a satisfactory account of what it is for him to grasp the thoughts or meanings we ordinarily ascribe to him. To read Dummett's equivalence thus, from right to left, in the case of either theories of language or theories of thought, is to be committed to providing a genuinely explanatory account of what it is to know certain meanings and to think certain thoughts. So we shall call someone who adopts this position an Explanatory Theorist

The less ambitious project for someone occupying the middle position is to describe what speakers know and what their having that knowledge consists in; where the latter part of this task is undertaken within a theory of knowledge for these speakers. Such an account must describe what it is for someone to have the knowledge the theory ascribes to him, given in

terms of a non-psychologistic, non-naturalised epistemology. We could call such a theorist a Descriptivist since he takes himself to be describing facts about speaker's knowledge of meaning. He may choose to give these descriptions in a full-blooded way, where he would be obliged to tell us what facts about a speaker he is describing, and why such facts about a speaker amounts to that person's having knowledge of a language of the sort characterised in the theory of meaning.

The least epistemologically constrained position is the avowedly modest proposal of the first theorist. He merely describes the meanings of someone's words and to that extent takes himself to be describing what is *known* or *knowable* by someone who can understand that language. It is natural enough to see any explanatory pretensions this theorist might have as being merely those of intending to *interpret* the speaker's language; so we could call him an Interpretation Theorist in Davidson's sense. This theorist would be an attributionist about speakers' knowledge of meaning. Any articulation of a speaker's knowledge would be a reflection of the structures exhibited in the theory of his language.

So we can think of these three readings of the equivalence thesis as providing three construals of the theory of meaning. We can think of them as three stances we can take towards such a theory: the Interpreter's Stance, the Descriptivist's Stance or the Explanatory Stance.

6.2) How Should We Construe the Theory of Meaning?

I take the question of how we should construe the theory of meaning as the question of which of the three stances on the above classification we should adopt towards the theory of meaning. Let us examine this now.

If we return to Dummett's own view and ask where he belongs on this classification, it seems natural to place him in the middle. He wants to give an account of linguistic meaning that describes genuine knowledge of the speaker; but his stress on the epistemological constraints on a notion of meaning do not extend to cognitive psychological constraints on it. Yet on the other hand Dummett does want to say what it is for a speaker to have the knowledge which a theory of meaning would ascribe to him; and he does want to propose some substantive conditions on what it is to

Chapter Six

possess that knowledge. Dummett sees such knowledge as consisting in the speaker's possession of a complex capacity to use the language, and he regards the theoretical account of meaning as a representation of this capacity. One important question for this position is whether it can account for the structure of our linguistic abilities. We must do more than just ascribe to speakers a complex ability to use the language that somehow issues in the various aspects of our linguistic practices. It seems that we want an articulation of this complex ability into smaller sub-abilities relating particular abilities to particular aspects of the language. But the question is whether the segregation of these individual linguistic abilities should reflect the structure the theory discerns in the language, or should more directly reflect facts about the structure of speakers' competence. And if the structure of particular linguistic abilities is owed not just to the theory that describes them, but to properties of speakers, then Dummett is obliged to say more about just which properties of a speaker they are.

At times Dummett seems to view axiomatisation as not just an articulation of the theory's structure but also as a reflection of the speaker's conceptual resources. He offers this picture as a contrast with Davidson's holism where the deductive organisation of the molecular truth theory can play "no genuine role in the account of what constitutes a speaker's mastery of his language" (Dummett 1975a, p116). This suggests that, in contrast with Davidson, Dummett is keen to ensure that the theory's carvings of sentences into parts and wholes should be aligned to the conceptual structure of a speaker's linguistic understanding. But on the other hand he believes that the way to find out what that conceptual structure is like is not to investigate the cognitive structuring of an individual language user but to provide a theory of meaning that describes linguistic use correctly. But now we need to know exactly how we are to describe the behaviour that constitutes language use. And it is here that we find in Dummett an ambivalence about being drawn into empirical theorising which leaves him oscillating between the second and the third positions. He believes there are empirical facts to be captured about the character of a speaker's knowledge while at the same time believing that this is not a matter for empirical investigation. The theorist of meaning can proceed without detailed empirical inquiries to propose conditions on

Chapter Six

language mastery which can then be checked empirically. These include ascribing to speakers' certain cognitive capacities such as those which enable them to observe the conditions for a statement's truth, or to detect the evidence that bears upon a statement's truth. These are genuine epistemological claims about what an incipient language user must know if he is said to understand what the competent speaker understands by these sentences; and since meanings cannot be divorced from understanding, these are genuine epistemological conditions on the meaning ascriptions we should go in for in an adequate theory of meaning. Thus for Dummett, we want a theory of meaning that says what speakers' actually know, but which does not require any empirical investigation of the actual cognitive capacities involved. He also seems to think that we can undertake this project "without first having undertaken any epistemological inquiry at all" (Dummett 1978, p89) But this is odd for one who believes that the epistemology of understanding is based on the human cognitive capacities that constrain what we can think, see and mean. Dummett's concern that we should respect these epistemological limits when constructing a theory of meaning seems to be motivated by his antirealism - seen here as the thesis that truth is epistemologically constrained. But at the same time, executing the task of constructing a theory of meaning is meant to serve as a ground for his antirealist conclusions. This is an unhappy state of affairs since it leaves Dummett's claims that a truth conditional theory of meaning should "explain meaning in terms of actual human capacities for the recognition of truth" (Dummett 1976, p136) as resting upon uncashed assumptions about our epistemological limits, and hence open to the criticism that McGinn levels at him that he is simply pre-supposing "a prior inventory of recognitional capacities" (McGinn 1981). The reply might be that the onus is on the realist to come up with some account of how a speaker could have acquired knowledge of the verification-transcending truth conditions of a statement, or succeed in manifesting his grasp of them to others. But all such challenges depend on background assumptions that there will always be a flaw in the realist's attempt to do this, which presupposes some uncontroversial view about just what can be acquired or manifested in the normal course of things. So just as any attempt to describe a pathology

assumes some model of normal functioning, what the antirealist needs to offer is some account of the types of content exercises of our recognitional capacities ordinarily deliver. For without further investigations what can Dummett offer in the face of an attributionist challenges to his account of what someone with these and other linguistic abilities knows?

The attributionist who is also a semantic realist may say that even in the basic cases of language use what the speaker is required to apprehend when observing a state of affairs necessary and sufficient for the objective truth of some observation statement is what it is for that state of affairs to obtain undetectably, or at other times. So even at this most basic level of language use a disagreement could arise between realist and antirealist as to how we should conceive a speaker's knowledge of such a statement's truth conditions and hence about what we should assume he can display to others of his knowledge. The upshot is that if there is a genuine dispute between realist and antirealist as to what contents we actually succeed in grasping and expressing, and, if Dummett believes there is a fact of the matter here, it won't do to settle the dispute by resort to the the linguistic abilities as they are represented in one's preferred theory of meaning, or to one's epistemologically preferred way of describing the behaviour manifesting our possession of those abilities. The crucial question is whether one theory or another accurately represents the contents of the understanding acquired, deployed and displayed by native speakers of that language. The search for epistemological conditions on meaning assignments, as Dummett sees them, may well have to be grounded in something else, and I submit that an important way to get at some of the essential features of the semantical saliences picked out by our network of recognitional capacities is to turn to the cognitive psychological underpinnings of these capacities as described by empirical theories of cognition. The success of this project will depend on how satisfactorily the empirical psychology provides a detailed treatment of many aspects of intelligent human behaviour originally identified in our ordinary, non-scientific way. Within the framework of such an account the realist would be obliged to explain how we arrive at a conception of certain states of affairs with which we are wholly unconnected by our cognitive powers. It is hard to see how such a naturalistic account of our knowledge could

make room for our representation of verification-transcendent states of affairs, whereas the antirealist position might put no strain on an account of our cognitive capacities grounded in our informational commerce and causal contact with the environment.

Meanwhile, the position Dummett is trying to occupy is that of a priori theorist who preserves the theory of meaning as a purely philosophical project while trying to insist, in opposition to the attributionist, that there are linguistic facts, and facts about speaker's knowledge, which are not merely artefacts of the truth-theory, or matters of external attribution. Pressed to say what these facts are, Dummett will say they are our abilities to use words as we do; but asked to say more about how we are able to use them he will claim that he is not describing the interior psychological states of the language user or attempting to "explain how we are able to use [a] word as we do", but that he is simply providing "part of an extended description of what that use consists in" (Dummett 1973, p681) But what does this use consist in exactly? All the observation of a speaker's use licences us to ascribe is a behavioural disposition of the speaker to continue using the word in the *same* way. But we have still to be given a full characterisation of what that 'same' way is and what content the expression so used is thereby revealed to have. Without this how can we know which specific ability a speaker is exercising and what knowledge this involves. All we observe of the facts of meaning in use is a wide range of behavioural practices, but how should we describe these? The attributionist will have one view of this, but what is the alternative of the descriptivist who claims to be faithful to the epistemology of speakers' understanding?

The only distance Dummett can gain from the bare facts of linguistic behaviour - themselves insufficient to support an account of linguistic meaning - depends on his claim to descry evidence in those behavioural events of a speaker's exercise of different abilities. But what is Dummett to say about those abilities? He must either treat them behaviouristically or mentalistically; either as merely externally displayed discriminatory skills, or as capacities that depend on the internal organisation and functioning of the speaker's cognitive apparatus. Of course neither of these options will supply every part of our account of the contents of a speaker's

Chapter Six

utterances, but the question is where we are to locate the facts about linguistic meaning and understanding. Whatever the full story about the contents of speech in a language might look like, can the theory's description of them be supported by the pieces of speaker's behaviour characterised without recourse to the language in question, or must they be taken to reside in those psychological states of the language user involved in the causal aetiology of his behaviour? Dummett wants to resist behaviourism, wants to maintain a full-bloodedness requirement, but does not want to go all the way towards the explanatory stance. But in doing so what we see is a hopeless oscillation between attributionism and some more committed explanatory position.

When resisting the attributionist, Dummett makes remarks about what it is to possess the theoretically described knowledge which tend in the direction of an explanatory account of speakers' linguistic capacities. He criticizes Davidson's account of knowledge that suffices to understand a language as being:

... somewhat roundabout unless [the] ability to speak a language actually does involve having such knowledge (Dummett, 1981(b), p5)

The trouble as Dummett sees it with Davidson's handling of the theory of meaning is that it does not:

... explain *how* the speaker is able to understand the language, ie. what renders him capable of doing so: it at most provides a somewhat oblique characterisation of his linguistic ability. (loc. cit)

This suggests that we need a more psychologically realistic, non-oblique characterisation. However, at other times, in his efforts to resist what he sees as psychologism in the theory of meaning, Dummett seems to revert to the position of the interpretation theorist who talks about giving an account of *what* a speaker knows, and not *how* he is able to know it:

A theory of meaning is not a rendering of what a speaker knows in knowing the language, but a systematisation of it. (Dummett 1986b, p150)

Chapter Six

A successful theory of meaning explains how the sentences of the language express thoughts by displaying what it is for them to have the meanings that they do. What does this is not just that part of the theory which specifies their meaning according to its own technical characterisation of meaning, but the theory as a whole, including the principles for deriving other features of sentences from their meaning as so characterised; for it thus exhibits the entire functioning of the language. (loc. cit)

In no case is a theory of meaning intended to give a representation of what in fact a speaker knows, any more than an axiomatisation of a mathematical theory hitherto unaxiomatised is meant to represent the previous state of that theory. (loc. cit)

Even though Dummett stresses that:

Only a theory of meaning that employs notions directly based, unlike those of truth and falsity, on features of linguistic practice can be reconciled to the character of our knowledge of our mother-tongue. (ibid., p165)

There are plenty of remarks by Dummett that give credence to one side or the other. I see Dummett's difficulties here as due to the inherent instability of the Descriptivist position he wants to occupy. There is no middle way between a priori attributionism and a commitment to describe genuinely explanatory facts.

This tension in Dummett's position surfaces mainly in his discussions of the structure of linguistic abilities, where he describes something like a psychological structure for them, only to dismiss the idea that our actual knowledge could really be psychologically structured in this way. For example, Dummett is prepared to say that:

... a knowledge of a language involves a knowledge of its syntax, and this requires a classification of words and phrases into syntactic categories, so that we may attribute to one who has the capacity to speak grammatically a tacit knowledge that a given word is, eg., a noun. (Dummett 1976, p71)

He goes on to suggest that:

Chapter Six

..the capacity to recognize certain sentences containing that word as well-formed, and others as ill-formed, depends upon knowing the syntactic categories of other words and the complex rules of sentence-formation which may be expressed in terms of those categories. Here, an implicit grasp of certain general principles, naturally represented by axioms of the theory, has issued in a capacity to recognise, for each sentence in a large, perhaps infinite range, whether or not it is well-formed, a capacity naturally represented as the tacit derivation of certain theorems of the theory. (loc.cit.)

However, when it comes to talk about the actual knowledge a speaker has, then Dummett's talk about knowledge of axioms is just a *façon de parler*. The only knowledge Dummett can credit a speaker with is knowledge of the theorems, and this is because it is only the theorems which correspond to recognisable, specific linguistic abilities of the speaker: in this case, abilities to recognise of any sentence whether or not it is grammatical. Exercises of these abilities are said to manifest his implicit knowledge of particular theorems. A speaker has *implicit knowledge* of a theorem when it provides a theoretical representation of a specific ability in which his knowledge is taken to consist. Dummett expresses his reservations about this notion when he says of a speaker, "we are not explaining, but merely *characterising*, his practical competence, his possession of which simply constitutes his [implicit] knowledge of the theory ". (Dummett 1983, p110). It is because he sees himself as 'characterising' the practical abilities shown in behaviour that he is stuck when it comes to implicit knowledge of the axioms. He feels the need to make use of them, to describe the deductive shape of the theory and to appeal unhesitatingly to the process of derivation, but because he can isolate no single ability in behaviour to manifest knowledge of an axiom he can characterise no separate ability as a component of the language user's knowledge, and so is forced to deny any psychological reality to separable abilities that constitute knowledge of words in the speaker's mind. But we might question Dummett's argument here. It seems to be that since possession of linguistic knowledge is constituted by possession of specific linguistic abilities, and no specific abilities corresponding to the axioms are identifiable in a speaker's behaviour, he cannot be accredited

Chapter Six

with separate knowledge of the axioms. But it is only because the abilities are not *behaviourally* isolable that speakers are not accorded knowledge of the axioms: but what is behaviourally inscrutable may be cognitively discriminable. But for this we would need recourse to a theory of cognition. Dummett resists this move but it is not obvious that Dummett wants to rely on the behaviourist premise either. But the truth is he has no other clear way to separate out particular abilities from mere observation of a speaker's language use. The best he can do is this:

... the ascription to the speaker of an implicit knowledge of those axioms is based on the confidence that he has a general capacity which embraces all the specific abilities which correspond to theorems derivable from that set of axioms. (loc. cit.)

But we get no further analysis of that general capacity except by the deductive structure of the theory, so we can begin to wonder whether this is a purely theoretical characterisation of structure like that offered by the attributionist. Moreover, no account is forthcoming about what it is that *issues in* these specific abilities. Were he not restricted to examinations of behaviour, he might find articulations in the cognitive states of speakers which mirrored the deductive structure of the theory, and could use these to give some substance to an articulation of our linguistic abilities. These could be dependent states, having no significance on their own, but which together enter into one's structured understanding of a language. Dummett simply refuses to look for such hidden details and rests content with an external description of linguistic use. But it is doubtful whether language use can also show us what gives rise to it, or what gives it the form and character we recognise in it.

In holding to an account of meaning in terms of a description of observable use, Dummett's makes the remark that the theory which represents a speaker's abilities:

... is not open to assessment in the same way as an ordinary empirical theory; it is not to be judged correct merely on the ground that it tallies satisfactorily with observed linguistic behaviour. (Dummett, 1975b, p15)

Chapter Six

Does this mean the only alternative to behaviourism is attributionism? It would appear so, just as long as one sticks to Dummett's misleading portrayal of how empirical linguistic theories can be tested. But the other alternative is cognitivism which I shall motivate below.

Notice that what holds for the case of syntax holds at the semantic level too. Dummett believes that the theory of meaning is finitely axiomatised in the form of a truth theory, with the notion of truth understood in terms of warranted assertibility. But he has no way of substantiating knowledge of the axioms here either, and so for Dummett, everything below the level of the sentence should be treated instrumentalistically. But this position sits ill with many other things he says, and in particular with his semantic molecularism where "a speaker of a language derives his understanding of any sentence of that language from his knowledge of the meanings of words" (*loc cit.*) If the meanings of words are just abstractions over the meanings of whole sentences, then we seem forced to treat this account of understanding instrumentalistically too. But now this undermines any attempt to justify the psychological reality of semantic structure and reference-determining word senses. And yet Dummett seems to follow Frege in thinking that :

... the understanding of the sentence, which is something cognitive, an act of mind, mediates between the semantic features of the sentence, in the strict sense, that is, its structure and the references of its constituents, and the actual employment of the sentence. (Dummett 1981a, p461)

In particular, a speaker's grasp of the truth conditions of sentences in which a proper name recurs is in each case a grasp of something which involves "knowledge of what has to be true of any given object for it to be the bearer of the name" (Dummett 1976, p136) If this is a particular piece of knowledge the speaker has and exercises on a number of occasions then his grasp of the meanings of individual sentences must be structured. And indeed Dummett says in "What is a Theory of Meaning" that the deductive shape of a theory of meaning: "makes due acknowledgement of the undoubted fact that a process of derivation of some kind is involved in the understanding of a sentence." (1975b, p112). Again and again

Chapter Six

Dummett seems torn between being merely descriptive and saying something more explanatory. What is he to do? Were he to stick with his non-psychological treatment of implicit knowledge where one is said to implicitly know the theoretical description that represents some practical ability just in case one has that ability, he will suffer all the problems of the attributionist. In particular, how is he to reply to the charge that not every way of representing what a person does correctly captures that person's knowledge. To be sure of describing how things are with the agent we must find a guarantee that he is acting intentionally under the description we give of him. In the case of linguistic action this amounts to finding a description of the speech act and content a speaker expresses by his utterance. But if this is just a matter of finding some rule-governed description of his behaviour, there will be many ways to do this. And to say that a person implicitly knows a rule just in case it represents the pattern of his unfolding ability to use some sentence or expression will mean that he implicitly knows any of a multitude of possible rules that could represent his patterns of activity. How are we to impose further conditions on when a theoretical representation correctly describes how things are with our subject? If we can find no such conditions but we persist with Dummett in thinking there is something in virtue of which a speaker is expressing one sense rather than another, then we are faced with the uncomfortable conclusion that not all of these ways represent some actual feature of the knowledge a speaker can be said to be displaying but that we cannot tell which one does. This is entirely unacceptable to Dummett since it would leave some aspect of meaning and speaker's knowledge hidden and incommunicable. And yet Dummett tells us that a speaker knowledge of meaning must be fully manifest in his linguistic behaviour. But given that a theory of meaning's fit with linguistic behaviour does not enable us to judge of its correctness, how are we to judge a theoretical representation of what a speaker knows? If a judgement here just consists in some theoretical decision this would tend to undermine Dummett's descriptivism and full-bloodedness about what it is to possess the concepts a speaker is said to be exercising in his competent use of the language. What is Dummett's alternative? Consider for instance Dummett's sketch of what it is to grasp the concept "square":

Chapter Six

At the very least, it is to be able to discriminate between things that are square and those that are not. Such an ability can be ascribed only to one who will, on occasion, treat square things differently from things that are not square; one way among many other possible ways, of doing this is to apply the word "square" to square things and not to others. (Dummett 1975b, p7)

There is nothing trivial or question-begging in such an account since although it makes use of the concept square it only uses it in first intention and not in the context of a that-clause describing the knowledge or thoughts of the subject. The troubling question is whether the account will do. Any number of finite manifestations of this discriminatory skill will not distinguish it as an exercise of the concept 'square' from the exercise of many disjunctive concepts of the form "square or....". This criticism of Dummett's account by John McDowell in "In Defense of Modesty" (McDowell 1987) is not relieved by turning to the subject's use of the word 'square' for we need know what concept it expresses for him. We seem bound to say that there are many different rules which could represent his ability to use the word: one for each of the disjunctive concepts whose conditions of application provide alternative rules for his use of the word. Must a subject be said to have implicit knowledge of all of these? Surely this would devalue the notion of a speaker's implicit knowledge of the meaning of his words. Perhaps the way ahead is to point out that while a speaker may fit many rule-governed descriptions it would be hopeless to say that he is guided by all of them. But which one of them does guide him? Of the rules someone could be said to be following Dummett says:

What these rules are is not open to immediate inspection: they do not for instance exhaust the observable regularities in play. (Ibid., p13-14)

No amount of inspection will rule out the alternatives. So Dummett seems forced either to say that we cannot know what a speaker implicitly knows about his language since he cannot manifest that knowledge in his behaviour, or to give up the idea that a speaker is actually following one

particular rule rather than another. The latter option might be open to Dummett for the following reason. A speaker *manifests* his knowledge in behaviour but his possession of that knowledge consists in possession of an ability which is only exercised but never fully revealed in that behaviour. This need not mean that there was some hidden component to a speaker's knowledge of meaning since the inference from observations of someone F-ing to his having the ability to F is not an inference from effect to cause. So we need not think of the knowledge-constituting ability in terms of some hidden component of a speaker's knowledge that *guides* his linguistic behaviour in a causally relevant psychological sense. But now by the same token it becomes utterly mysterious how we *are* to characterise the speaker's ability to use words in a non-question-begging way; and difficult too, to see how we are to account for the speaker using those words in the same way over time. At this point the Descriptivist faces the option of moving wholeheartedly towards attributionism. But of course this offers no non-question begging way of saying what a subject knows, and in some cases, no way of saying what he knows at all. Furthermore, if a subject's behaviour is susceptible to descriptions in terms of particular semantical regularities then we will want some account of what renders it thus amenable to these descriptions; ie. we shall want some account of what its exhibiting those very semantic patterns consists in.

I suggest the answer to McDowell's underdetermination objection might lie in adopting a more naturalistic and less a prioristic account of concept-mastery which would suggest that the acquisitions of the more outre disjunctive concepts is less evolutionarily likely for creatures like us, functioning well in our environment, than is our acquisition and stable possession of the simpler concepts we could make use of. It is simply more likely that creatures with our cognitive functions, when acquiring a concept by repeated exposure to square things, will acquire the concept square, since it is more likely to be the normal function of the state which produces an output for a perceptual input of something square, or the ability that is exercised in response to square things, to indicate the presence of something square. This appeal to simplicity does not involve speakers in inductive hypothesising about the meanings of others words,

or in comparing the idiolectic character of their own understanding with that of others. So it does not risk the pernicious psychologism McDowell envisages, which would make another's meaning a matter of guesswork as to how things stood in some concealed, private sphere. The concepts acquired by, and recognised in, the behavioural manifestations of others would be those which it was most natural for cognitive systems with our information-gathering facilities to acquire and recognise.

On a related issue, Dummett's inability to offer anything for the theorist of meaning to be describing accurately does nothing to allay the fears that the Descriptivist has nothing non-attributive to say about why theorists describe the contents of speaker's sentences as structured. Dummett is in a worse position than the attributionist, so far as his claim to be describing speaker's implicit knowledge of semantically structured truth conditions goes, since he cannot just claim this structure is an artefact of theory as the attributionist can. And finally, attributionism should be resisted if one wants to retain, as Dummett does, the idea that a speaker's use of his language is guided by his knowledge of language. But how can Dummett sustain this? The only satisfactory option I can dimly discern here would have to involve the thought that the guiding mechanism of linguistic use was not itself a mental notion (or indeed content-involving) but that it was a causal mechanism responsible for a behavioural disposition which got subsequently re-described as a genuine linguistic ability when the unfolding pattern of word use created by exercises of that disposition inherited its identity in the setting of a communal language practice. This might require the following set up:

The verbal output provides the "pattern" for assigning semantic properties (meaning or content) to those internal, neurophysiological states that produced it. The internal structure of our cognitive states is merely a reflection of the semantic properties of the output they produce. (Dretske 1984, p283)

But what is also required is some cogent account of how the communal practice of public language use succeeds in conferring semantic properties on the patterns there displayed. I confess I find it utterly mysterious how this is supposed to happen, unless by reference to some social psychology,

but then such a molar psychology would have to resist attempts to ask what enables individuals to participate in the practices thereby established; and it would have to do this by answering those questions itself. But nothing is forthcoming about what such an account might look like and how it might work. Talk of "forms of life" is the best we can do here. But since not every creature can participate in those practices, surely we want to know what properties of those who are able to do so equips them for this. And surely at that point we are going to have to appeal to facts about individual participants? Language use may depend on a society but it is a society of minds.

I suggest the only satisfactory remedy is for Dummett is to become more full-blooded and explanatory and be prepared to look for genuinely substantive conditions on what it is to possess concepts and to have a structured understanding of one's language. It has yet to be said how this can be done. However, it should be becoming clear that the real dispute between attributionist and non-attributionist conceptions of mind and language is whether we should adopt an Interpretive or an Explanatory Stance towards the theory of meaning; so far the middle position seems untenable. I shall now show why Dummett cannot adopt the former.

6.3) Not a Theory of Meaning, not a Theory of Understanding

In the light of the foregoing discussion we can focus on two earlier objections I made to Davidson's semantic programme. He offers an account of thought in terms of language and his theory of language leaves the content, if not the structure, of thought and language massively indeterminate. From the Explanatory Stance it is not hard to see why, since he cannot allude to the facts of the matter which determine how things are with the speaker. The main objections, as mentioned above, are that he cannot account for what speakers actually know, and that he credits them with more than they know. So he fails to satisfy both sides of Dummett's equivalence thesis. Let us look at these in order.

Since Davidson thinks that language is our only route to the character and content of someone's thinking, and since the question of which objects a speaker is referring to has no answer, then the question of which

Chapter Six

object he is thinking about has no answer either. The notion of individual belief as identified linguistically will be referentially inscrutable and plotted by theory.

It is at this point that Fregeans and Russellians on this side of the Atlantic begin to protest (although John McDowell wants to owe allegiance to Frege, Russell and Davidson). Many British philosophers have wanted to say that we cannot know what a speaker means unless we know which things the singular terms in his sentences are about. Some have even said that we cannot ourselves entertain singular thoughts unless there is an object for those thoughts to be about. The knowledge of reference is either explained in a reference-determining way (by Fregeans) or a reference-dependent way (by Russellians). Note that neo-Fregeans combine these accounts and end up with object-dependent senses.

Further elaboration of these views is needed, but it is useful to point out straight away that these requirements are taken to be conditions on linguistic understanding, where the account of understanding is an account of the meanings known to the speakers of the relevant language. So what we have here is a strong repudiation of Davidson's conception of an interpretation as an account of linguistic meaning or understanding on the grounds that there is more to be described than this.

For Dummett, understanding a sentence containing a singular term involves knowing what it is for something to be the referent of that term. For concrete terms this might, in some cases, amount to the subject's possession of an ability to recognise an object as the bearer of that name when appropriately presented with it. However, there may be a variety of ways to explain our knowledge of the identity of a term's reference. Whatever mode of identification is given it is provided in support of a general principle that to understand a sentence containing a singular term one must know which objects and properties the terms and predicates of that sentence concern. Only thus can one know what another person asserts or denies, for only then can one know what it is for *what is said* to be the case.

Gareth Evans would go so far as to contend that one does not understand *what is said* by an utterance of a sentence containing indexical or demonstrative expressions unless one knows enough to identify the

Chapter Six

references of those expressions. This is clearly to ask for more from the notion of meaning than Davidson thinks is possible. While to understand only what Davidson suggests an interpreter can recover of meaning on occasions is not enough *to be able* to understand the sentence used, according to Evans. So if Evans can substantiate his account of speaker's knowledge of meaning, he will have shown that Davidson cannot provide an account of what a speaker knows.

Davidson is not, of course, oblivious to the importance of indexicals and demonstratives in the everyday use of a language, and he even suggests that the application of a truth predicate should be relativised to speakers and times of utterance. Davidsonians then talk of harmlessly abstracting from these details when discussing theories of truth and interpretation more generally. But a moment's thought should tell us how odd Davidson's concession really is. Firstly, the relativisings will still not guarantee to the object language terms a local reference to salient objects and events; and secondly, how do these words in the language of the theory manage to refer when Davidson has argued quite generally that nothing about a language can show how its singular terms and predicates map onto objects. He even says, "I find it impossible to formulate the relativised concept of reference in an acceptable way" (Davidson 1984, p227). There is no mystery about how an interpreter can set up arbitrary alternative *reference schemes*, which match different object language terms and predicates with terms and predicates in his own language. But that is not to the point here. The relativisations to speaker and time are meant to do *real* semantical work. Relative to a language they can *specify* a reference, but these specifications have no empirical import, the terms make no genuine reference.

For Evans, this position would leave us hopelessly unconnected to the world and unable to think the thoughts required to understand certain basic types of speech. According to Evans, one part of what we need for a satisfactory account of demonstratives is to see the understander as enjoying the right informational connections with his environment; where this does not mean his merely thinking the thought, "I am connected to the object from which I am currently receiving this

Chapter Six

information", which could not bridge the gap. It requires him to be *in* those informational relationships with things.

More generally, Evans supposes that "a subject cannot make a judgement about something unless he knows which object his judgement is about" (Evans 1982, p89). The subject must be able to distinguish the something he is thinking of from all other things. Evans calls this Russell's Principle (a liberated descendant of Russell's Principle of Acquaintance). What is needed is an account of what it is for the subject to *know which* object he is thinking or speaking about, where "knowing which" provides discriminating knowledge of that object. To do this, Evans is required to do two things: (i) to give an account of the thinker's capacity to distinguish that object from all other things, which amounts to his having identifying knowledge of it, and (ii) to say why thoughts about individual particulars requires the subject to be able to do this. The former is given in terms of different modes of identification that enter into our thinking about particulars. These are: the descriptive, the demonstrative and the recognition-based identifications of individual particulars. For the second part of the requirement Evans offers a theoretical defence of Russell's principle that unifies these different relations in which thinking subjects stand to objects. The account Evans then gives is nothing short of what he takes to be the correct conception of singular reference and its role in conceptual thinking and the ascription of thoughts. The defence is tied to an account of how perception delivers thoughts about particular objects and how concept-exercising and reasoning is only possible atop "an informational system - which constitutes the substratum of our cognitive lives". The explanatory work for a theory of thinking, that explains our ability to think about material objects (non-descriptively), is to postulate connections between the two cognitive levels: the primitive and the more sophisticated level. For this we need some theory of the levels of cognition that would identify the primitive and the sophisticated layers of our cognitive lives. This is the background to the theoretical defence of Russell's Principle. The defence itself amounts to a lengthy discussion of the inadequacy of alternative accounts of reference that deny the Principle. But to someone like Davidson who denies any substance to the relation of reference in the first place, this is a non-issue. The importance of all this

for our consideration of Davidson is that Evans' picture requires different levels of cognition and a theory of these, as well as introducing "the notion of *being in an informational state with such-and -such a content* as a primitive for philosophy"(ibid. ,p123). This is a belief-independent state which can embody non-conceptual content - such as proprioceptive information about one's orientation in behavioural space, as sitting up, falling over backwards, in front, behind, etc.. This is a kind of information Davidson either ignores or re-presents in conceptual terms by way of subject's beliefs. Either way, Davidson would see it as having no role in our theorising about the mental. Theorising of the sort Evans goes in for is merely "changing the subject" so far as Davidson is concerned. Given how little Davidson's attributed notion of mind recovers of the properties of mind and language, perhaps it is time we changed the subject. However, the serious objection would be that the attributionist has just appropriated the concept of mind. Evans' rebuke to Davidson consists in showing that nothing less than the postulation of a layered theory of cognition is needed to give an adequate account of our cognitive lives, which for Evans includes perception, memory, communication, and knowledge. He claims that it is in these underlying levels of cognition that certain essential features of our cognitive and linguistic abilities inhere. This is a large project, of which I hope to tackle just one small corner: the need for cognitive underpinnings in an account of our knowledge of language.

Without rehearsing the battery of arguments for the existence of a genuine notion of reference, including the argument that there could be no notion of sense without it, let me just record my agreement with Frege, Dummett, Evans, and Wright (1983, p78ff), that an account of meaning sufficient to capture the facts about a speaker's linguistic understanding of a sentences containing singular terms needs to include an account of what it is for individual particulars to be *knowably* the references of those terms. My strategy will be to accept that the sense of a linguistic expression is always given in terms of our way of thinking of, or being presented with, the reference of the expression; and that if the notion of sense can be made good by giving it some psychological reality, then there will be some substance to the notion of reference, pace Davidson. Although I accept the idea that linguistic sense involves the notion of linguistic reference, I do

not accept Dummett's claim that sense as a constituent of thought always has to be the sense of some linguistic expression or other. A variety of recognitional abilities can offer examples of reference-dependent, or reference-determining, senses which preserve the intentionality of thought in a non-linguistic medium. However, I also believe that the kinds of recognitional abilities we acquire, and the set we develop as permanent features of our cognitive make-up depend to a large extent on the language we speak and the linguistic community in which we learn our mother-tongue. Now if the senses speakers attach to their words and sentences are part of what they have to know to grasp the meanings of those sentences, then a theory of meaning will have to treat knowledge of what it is for objects to be the references of certain words to be a necessary part of speakers' understanding and ability to use their language. Without it, neither the speakers themselves, nor the meaning-theorist, would know what is was for what is said by their sentences to be true. The knowledge of what it is for something to be the reference of a singular term occurring in a sentence must be part of the linguistic knowledge recorded in any theory of meaning which says what the speaker understands by that sentence. So it is an adequacy constraint on a theory of meaning that it should assign references to a speaker's terms compatibly with a model of this knowledge. Now if knowledge about the references of constituents enters a speaker's conception of a sentence's truth conditions so will its semantic structure. So we could require of any theory of meaning for a speaker's language which provided a correct description of his understanding that it should assign the correct structures and references to his sentences and their parts. Structure and reference constraints on theories of meaning would cut down vastly on indeterminacy. But they can only serve as genuine constraints if they can be related to aspects of speaker's knowledge. If this proves possible then the basis for a correct identification of someone's language will be what Davidson's theoretical descriptions deny us: what speakers know of their own language. The challenge is to substantiate this objection in the face of Davidson's competing attributionist picture of the mind of the language user.

Chapter Six

It is time to turn to the second objection to Davidson's approach: that he credits speakers with more than they know. From the interpreter's stance, meaning is a matter of the construction an interpreter can put on someone's words. For this he needs a theory that extracts the meaning in the context of neighbouring meanings for sentences and supporting propositional attitudes of the speaker. The theory has to isolate those parts of a speaker's behaviour which count as the bearers of particular semantic roles within sentences, and semantic roles of sentences within the language. No speaker or interpreter commands a clear sight of the whole of a language, but he is supposed, instead, to have command of an evolving theory which he adjusts to the demands of interpreting anyone he encounters:

In practice an interpreter keeps the conversation going by adjusting his theory on the spot. The principles of such inventive accommodation are not themselves reducible to theory, involving as they do nothing less than our skills at theory construction. (Davidson 1984, pxix-xx)

But why should the principles of conversation be required to account for linguistic understanding? I submit there is a perfectly intelligible sense in which you can read and understand the words on this page and you do so without need of an overall theory of me. You may wonder why I have said what I said, but you are in no doubt about what I said. A theory of meaning has every right to set out to be a systematic theory of that notion of linguistic understanding. The philosophically challenging task will be to say what facts about a speaker such a theory is describing.

But for Davidson, since meaning is plotted by a theory that interprets a vast range of linguistic and non-linguistic behaviour, its theoretical status leads to the implausible suggestion that the only way a language user can know the meaning of someone's words is to be in possession of a theory of the language to which they belong. This is implausible for three reasons. Firstly, there is precious little evidence that speakers know any such things as a truth theory. Secondly, if they did, there is a problem about what their knowledge of the language in which the theory is couched consists in.

Thirdly, as Dummett points out, there must be a way of understanding a sentence which is not a matter of putting an interpretation on it.

How would Davidson reply? His only route as far as I can see is to take the line taken in the case of psychological interpretation, where one could distinguish between the subject, the interpreter and the theory of what the interpreter knows. Can one be an interpreter of language without being a theorist of linguistic interpretation? Davidson talks of regarding the interpreter "as if he were using the theory we use to describe his competence", but how plausible is it to treat communication as if it were interpretive theorising? Well, Davidson could say that when we communicate we do not share a common thought, or obey common conventions or rules: we simply appropriate people's words and put the best interpretation on them we can. We must always understand others in our own terms, and so how reflective or theoretical the process of interpretation is will depend on the degree of difficulty involved and the closeness of someone else's language to our own. But the attributionist is supposed to think that all that is available to an interpreter is all there is to meaning. So, talk of closeness of meaning or structure would be strictly misleading here.

Davidson's need to turn speakers into interpreters is partly due to the lack of intersubjective synonymy between speakers: we cannot presume any two speakers share the same holistic understanding of the language they both speak. But to say that speakers re-interpret others in terms of the meanings they use does not explain a speaker/interpreter's knowledge of his own language. If we try to side-step this problem by simply providing a theory of meaning (interpretation) for the interpreter's language, this would not show that it was a theory of what that interpreter himself understood. Does this not put a strain on the relations between meaning and mind? This depends on one's view of things. Davidson could insist that the meanings assigned to a speaker's words must always comport with the beliefs we can ascribe to him. Furthermore, the account of the meanings of someone else's language has to be an account of the meanings that any well-informed interpreter could know. So a connection between meanings and minds is preserved. But to some this connection will seem highly problematical, for the theory of meaning is now to be

Chapter Six

regarded as a theory of what the interpreter understands: meanings are objects of the interpreter's knowledge. So the theory of meaning for a language is a theoretical account of what every interpreter knows, not what every speaker knows. The ground-floor level of linguistic comprehension is what is missing on this account, but this is a straightforward consequence of Davidson's third-personal perspective on the linguistically revealed mind. From the interpreter's perspective one is always making sense of the speech of others by ascribing beliefs, desires and meanings which portray their psychological life. But now we want an account of the interpreter's understanding of the materials he uses to give these interpretations. How does he understand the language in which the interpretation is expressed? To answer this, we are required to ascend to the next level and provide a theory of meaning for the interpreter's language. But this will not give us the interpreter's own understanding of this language. We are faced with a dilemma: if the interpreter's knowledge of his own words depended on his grasping of the theory we would be faced with a regress, and if we do not expect the interpreter to know such a theory, then knowledge of the meanings of his own words seems to be beyond the interpreter's reach. The second horn may seem unsatisfactory but it is this move which successfully avoids the threatened regress. Instead of having more and more interpretations, each one standing behind the other, and each one having to be a genuine item of knowledge, the chain is broken by making the theory for interpreting a speaker/interpreter's language a *possible* object of human judgement as opposed to an *actual* item of knowledge. A speaker is not required to know a theory for his language but he has to *satisfy* one. There must be a theory which could, in principle, be provided for him. This theory must be potentially knowable by other human subjects to be a theory of understanding. But since no-one need fill this role there is no risk of the regress. The fact that two speakers in a conversation are making sense is to be explained by the possibility of a third-person making sense of what they say by means of the theory. They make sense only because of the possibility of interpreting them, and because of the possibility of their being able to master the theory of their conversation.

Chapter Six

There are a number of difficulties with this position. For a start, it can provide no satisfactory account of a speaker's understanding. Davidson's suggestion would be that a speaker could not be said to understand his own language unless he was *capable* of grasping a theory for it. This idea involves no contentious claims about a speaker's actually having theoretical knowledge but it does not adequately characterise the speaker's knowledge, for I agree with Strawson that a speaker's *potential* for grasping a theory cannot explain his *actual* mastery of his language and his capacity to engage in meaningful speech. (Strawson, 1966, p193). So we are still none the wiser about the interpreter's grasp of his materials for giving interpretations. And without this, there is an epistemological risk that all we are giving is *an* account of the meanings of speakers' words without providing the right account. There is just no notion of what a speaker's knowledge consists in over and above an interpreter's account of what a speaker's words mean and what beliefs can be ascribed to him. But if there is something knowable to recover about a speaker's state of understanding then Davidson's claims for the exhaustive potential of the methods of interpretation are false, and moreover, if that knowledge contains information that helps us to determine more of the properties of the speaker's language, Davidson risks misrepresenting the character of the language.

Worst of all, the theory of interpretation which a speaker satisfies, and in virtue of which he is a meaningful speaker, is a theory of his linguistic behaviour. This is the wrong level of description on which to append some of the important aspects of structure in a language. Syntactically, this picture is hopeless, for a theory of syntax, whatever it is, certainly isn't a theory of behaviour. Furthermore, an account which re-describes behaviour in intentional terms cannot provide any account of our capacity for linguistic production. The contrast would have to be with a notion of speaker's knowledge which stated information used in both comprehension and production, and neutral with respect to both. Talk of regarding the interpreter "as if he were using the theory we use to describe his competence" will not help with any of these objections.

Is there an alternative? One alternative would be to work out a closer link between semantic theory and a theory of cognition for speakers.

Cognitive theory characterises the internal psychological organisation of speakers; whereas semantic theory characterises the structures and semantic rules for sentences and expressions in a language. Prima facie, the two projects might be linked if the structured linguistic abilities of a speaker - those that determine the semantic properties of his language - turn out to be psychologically structured abilities. The notion that shows how a theory of meaning contributes to psychological explanation is notion of tacit knowledge, for the facts about cognition on which semantical facts could be said to depend would be facts about a speaker's tacit knowledge of semantic theory. This is not an account of the conscious propositional knowledge the ordinary speaker has, for he need not know any of the details of the theory in this way; instead it is a matter of certain properties which are true of him, such as his being in certain cognitive states whose contents are correctly described by the clauses of the theory. I shall discuss this notion more fully in the next two chapters. But before going on, let me end by suggesting why both Dummett and Wright, despite their resistance to such a notion (Dummett 1983, pp110-2, 1986b pp147-8; Wright 1981, 1987, pp204-238), need to embrace something of this sort to provide an alternative to Davidsonian attributionism. I believe that both of them feel the need to provide a more substantive notion of speakers linguistic understanding, but without fully diagnosing the need for a cognitivist theory of understanding which this entails.

6.4) Linguistic Abilities as Cognitive Psychological Abilities

Antirealists like Dummett and Wright are often accused of behaviourism because of "the fundamental anti-realist thesis that we have understanding only of concepts of which we can distinctively manifest our understanding" (Wright 1980, p221). But the charge of behaviourism confuses what it is to *possess* an understanding with what it is to *manifest* it. To possess linguistic understanding is to possess an ability of some sort; to manifest it is to exercise that ability. Thus the behaviour which counts as an exercise of an ability to use a particular word correctly can count as a

way of manifesting one's possession of the concept it expresses. The behaviourist's mistake is quite general: it is to confuse the evidence for a mental state with the state itself. To avoid this, the antirealist must find a way to separate the ability to use a given word or sentence from the behavioural evidence for that ability. But how is he to do this? After all, Dummett talks of an ability being exhaustively manifest in use. So if a linguistic ability is something over and above a speaker's behavioural use, we need to know how we are to characterise a person's linguistic abilities. This is just the difficulty we ran into above. For instance, are these cognitive or behavioural aspects of speakers? It would be hopeless to characterise them in purely behavioural terms since that would threaten behaviourism once more, and worse still, it would leave us with nothing informative to say about just which bits of linguistic knowledge speakers had succeeded in manifesting (cf. the banality of saying that a speaker succeeded in manifesting his own behaviour). It seems clear that if we are going to consider behavioural manifestations as evidence of the knowledge-involving abilities a speaker is exercising, we cannot characterise those abilities in behavioural terms alone. What we should say then is that a speaker's understanding consists in his possession of certain *cognitive* abilities which can be *behaviourally* manifested. However, it is all very well to say that our linguistic abilities have both a mentalistic and a behavioural side, but we need some way of substantiating the claim for a cognitive component in these practical abilities. Both Dummett and Wright appreciate the need for such a component but neither of them offers any substantial account of what it could consist in, nor sufficient evidence for postulating it. They could resort to seeing it as a purely theoretical posit, but to do so is to lose the advantage over the attributionist of being able to substantiate the notion of a speaker's structured understanding of his language. This is a lacuna which I claim can only be filled by a substantive and cognitivist account of our linguistic abilities. Let us look at the cases.

We saw above how McDowell criticized the manifestation requirement that someone's understanding should be fully manifest in behaviour characterised without appeal to the concepts of the language in question. His objection is framed in the form of a dilemma. Either the

understanding could not be fully manifested in behaviour, since the theorist's choice of which concept to describe the speaker as exercising would always be underdetermined by the available evidence; or it could be fully manifest, but only in behaviour characterised by prior appeal to the contents and concepts we suppose him to be expressing. McDowell's defence of modesty rests upon the claim that without presupposing a prior understanding of a subject's language we could not see his behaviour as the joint exercise of a number of specific linguistic abilities in which his knowledge of language consists. There is just too little in the behavioural evidence of linguistic use, as the antirealist conceives it, to justify the ascription of specific linguistic abilities and hence particular pieces of knowledge. With that said, McDowell now sees himself to be free to describe a speaker's linguistic competence as he chooses, and in particular, to be free to ascribe to him knowledge of realist truth conditions.

Wright's response to this begins by admitting that a speaker's linguistic abilities are only partially separable in their behavioural manifestations. Just as a belief is at the service of many different projects and can be manifested by many different behaviours depending on the background beliefs and desires, so a particular linguistic ability admits of numerous behavioural manifestations, and cannot be correlated with a any determinate range of behaviours. Wright suggests that this does not blunt the manifestation challenge, for the antirealist's point concerns the connection between understanding and ability. If a speaker displays his understanding in exercises of recognitional and inferential abilities then attributions of realist understanding of a sentence must be accompanied by an indication of the *abilities* which would manifest that knowledge. It matters not that that this ability can be behaviourally displayed in multiple ways; the realist/ antirealist dispute takes place at one remove from behaviour. Wright suggests that the realist can come up with no ability whose multifarious exercises could manifest grasp of evidence-transcendent truth conditions. But this strategy of inserting a wedge between the complex set of practical abilities that make up understanding and the holism of behavioural manifestation depends on finding some way to substantiate the postulation of stable aspects of a speaker's mind that are more than just the sum of their exercises in behaviour. They

cannot be identified with behavioural dispositions or else we face McDowell's dilemma once more. If they are just theoretical posits we fare no better than the modest attributionist. McDowell requires Wright to justify his identifications of these stable and projectible cognitive abilities without merely presupposing the contents and concepts they are supposed to confer on a speaker. It is not obvious how Wright is to do this. The abilities themselves seem more like the beliefs which admit of such varied manifestations, and indeed Wright suggests that a speaker's exercise of recognitional abilities are accompanied by the formation of beliefs that things are as they are recognised to be. So, for example, an exercise of our ability to use the word "red", which depends on our capacity to recognise red things, will always give rise to a belief in the speaker that the object he is attending to is red. But in what does the presence of this belief consist? It cannot be manifested in behaviour in any non-question-begging way because of the evidential holism Wright describes; nor can it just be a matter of attribution if there is a fact of the matter about exactly which concept a speaker is exercising. If Wright needs to appeal to beliefs to ensure the intentional significance of a speakers' linguistic abilities and identify the abilities actually manifested, then we need to know on what basis subjects are credited with those beliefs. The identificatory role Wright conceives for belief-attributions shows the need for some further condition on behaviour to warrant the claim that a speaker is exercising a genuinely mental ability. But since the precise contents of our beliefs are identified linguistically, we can hardly appeal to these beliefs to identify the specific abilities which help us to identify the content of people's speech. So as far as finding identity conditions for our linguistic abilities is concerned, the unexplained appeal to belief does not really help. Once again we face a dilemma. Either some non-linguistic theory of concept mastery has to be provided, or the auxiliary beliefs will have to be identified by taking an interpreter's stance to the speaker's behaviour like that taken by the attributionist. The latter could not satisfy the antirealist's concern to describe what speakers actually understand and the abilities they exercise and manifest in their linguistic behaviour. The former option already embroils the theorist in giving a substantive account of the

Chapter Six

cognitive structuring of a person's mind, as we shall see in the next chapter.

Turning to Dummett, it is interesting to note that his ambivalence as between descriptive and explanatory theories shows up in his discussion of the nature of the practical capacities we manifest in behaviour, and by whose exercise we manifest our linguistic knowledge. He tells us that any practical ability will have a theoretical as well as a practical component:

The skill of a ballet dancer, for example, involves not only the ability to perform certain movements of which those not so trained are incapable - the practical component - but also a knowledge of which movements to perform, which is the theoretical component; someone who was a natural ballet dancer in the sense that he could, without any training, execute fouettes and the like as soon as he was shown them would, until he was shown the steps, possess *only the practical component of the ability and not the ability itself*. (Dummett 1983, p111 italics mine)

But how are we to characterise the theoretical component in the complex ability to speak a language? If not by appeal to behaviour or conscious knowledge of the reflective language user, then by what? It is at this point that we would surely hope to appeal to the underlying cognitive states and processes that give rise to speakers' conscious judgements about their language. Our conscious knowledge of language may not be inferential but if there were underlying psychological states, which were not beliefs, but which were "part of the process leading to belief formation" (Stich 1978 p501), and some complex mechanism mediating between acoustic inputs and outputs in conscious linguistic judgements about the properties of a sentence, then we could describe the process of comprehension in terms of inference-like computations involving subdoxastic states that made no showing in consciousness. These low-level information states, the mediating computations, and our conscious awareness of language, would all be involved in the provision of an overall account of the theoretical component of our linguistic abilities. This would be an account drawing on conceptual and non-conceptual states, and it would commit us to seeing the notions of a speaker's

Chapter Six

knowledge of language as a complex property spanning more than one level of cognition. And this is precisely what we need.

Furthermore if we take care to distinguish this form of cognitivism from certain pernicious theses of psychologism then we will avoid much bad argument. Among other things, psychologism meant: that the objects of enquiry were mental images or subjective ideas; that the methodology committed one to introspectionist psychology; that the contents of mind conceived in this way were private and incommunicable; and that processes of thinking affects the products of thought. Cognitivism, on the other hand is the meta-theory of scientific cognitive psychology: its results and methods are open to public examination; it aims to provide testable theories of the individual that generalise across all individuals; it posits fields of representational structures in computational systems whose contents may depend on relations to the environment, as in theories of visual processing. What it shares with the doctrine of psychologism and what leaves it open to attack from anti-psychologists is that it conforms to the thesis that the theory of processing is relevant to the account of products. In that sense it challenges the sovereignty of a priori theorising in analytical philosophy, mixing empirical and conceptual issues. It may be for this reason that descriptivists like Dummett and Wright resist it, although they both feel the need to provide a more substantive account of our cognitive and linguistic abilities. The temptations of the interpreter's stance and the a priori isolationism still beckon. I shall now show why that resting place is unsatisfactory.

The problem for the descriptivist begins when we ask what do we grasp when we grasp a sentence's truth conditions. And as far as the explanatory theorist is concerned the most acute difficulties arise for Dummett when he tries to explain a speaker's knowledge of the truth conditions of basic statements. A non-basic statement is one for which it is possible to give a non-trivial answer to the question of what renders that sentence true. In these cases we already have an explanation of what a speaker must know if he is said to know such a statement's truth conditions. A basic statement - prototypically an observation statement - is one whose truth is neither a matter of inference from the truths of other statements, nor a matter of reductive analysis on the basis of truths of some other kind. The

Chapter Six

important thing about these cases, as Dummett stresses, is that although the meaning theory can give no non-trivial account of that which renders those sentences true, it must give a non-trivial account of speakers' *knowledge* of what renders them true. It does this by ascribing to speakers a capacity to *recognise*, in favourable circumstances, the condition for a sentence of this kind to be true. Since these recognitional capacities embody knowledge of truth conditions which is neither inferential or reductive, the meaning theory must attribute to speakers "a faculty of unmediated recognition". This confers on those who are favourably placed a direct access to those truth conditions. Dummett goes on to say of these direct recognitional capacities that:

..neither the speaker nor the meaning-theorist can say whereby he recognises the condition as obtaining" (Dummett 1982, p106)

But surely something more must be said if we are to know anything of the character of a speaker's understanding. Dummett's denial of this seems to rest on the thought that further investigation of the faculty of recognition is a task for the theorist of cognition, and thus not strictly relevant to the project of the meaning theorist. But I shall now argue that this is false. Unless we say more about such capacities to recognise truth there will be no knowing for sure what someone grasps as the meaning of those sentences, and in particular, there will be no way of knowing whether he grasps the structure the theory of meaning ascribes to such sentences. Thus there will be no way to connect the detailed articulations in the theory of meaning to the facts about actual speakers' understanding. Furthermore, an investigation of this faculty of recognition will give us the clearest idea of the semantic content at the base of our language. For the epistemological function of this cognitive component of a speaker's mind is to track the semantically salient features of the environment, the evidential circumstances which warrant us in asserting the basic sentences of the language; so it only by understanding what is involved in the exercise of these cognitive capacities, that we can know what sorts of things are being indicated, and thus what the semantically salient features of the environment are. I shall have less to say about this last point for I want to concentrate on the argument that language mastery requires

Chapter Six

structure in our recognitional capacities, to be explained by treating them as psychologically structured abilities.

Despite having nothing informative to say about what make a basic sentence true, the meaning theory will still provide articulations of the semantic structures of these sentences. Now if these semantic structures characterise part of a speaker's understanding of each individual sentence, then there must be a structure to his recognitional abilities. The question is how are we to account for the structure of a recognitional ability? This is important, for if speakers really do grasp the structure of their sentences, and it is only via the structure of their recognitional capacities that we can tell which structures they grasp, then it is only by investigating those recognitional abilities that we can discover whether a theory of meaning that assigns a particular set of semantic structures can serve as a theory of understanding for their language. The claim that it is only via the structure of their recognitional capacities that we can discern the structures speakers grasp for their sentences is a simple consequence of the fact that for these basic sentences a speaker's understanding simply consists in the possession of a recognitional capacity. So that much is clear. But as yet we have no reason to assume that speakers actually grasp the details of the theoretical articulations the meaning theory provides for their sentences. So at this stage we have a conditional claim: if grasp of the meaning of a basic statement involves grasp of its structure, then that structure must be part of the structure of a recognitional capacity. Arguing by detachment I will then try to show that the relevant notion of structure in a recognitional ability is that of its cognitive psychological structure. This will provide an argument for the relevance of the theory of cognition to the semantical project of constructing a theory of meaning for a speaker's language. It will show that the meaning theorist must advert to a theory of cognition to get at the actual semantical details of a speaker's language.

Let us begin with the conditional claim. Does Dummett endorse the antecedent: that grasp of meaning involves grasp of structure? It might seem that in the case of basic sentences, and atomic sentences more generally, Dummett does not want to attach any weight to claims for the speakers' knowledge of their sub-sentential structure or the words they

Chapter Six

contain. So, for example, he describes our understanding of words in terms of our understanding of sentences:

..what constitutes the speaker's attaching a particular sense to the word is his *using* the word - more exactly, using sentences containing that word - in a particular way. (Dummett 1981a, p52)

These ways of using words must be described without taking advantage of the prior grasp of their senses by a speaker; for we are trying to say why his using a word in this way constitutes his attaching a particular sense to it. So it seems that we must describe his use of whole sentences without 'taking advantage of his prior grasp' of the senses of the words they contain. But now it will be hard to see how that use can show his grasp of particular senses for each of the words involved. Perhaps this means that all a speaker can show is a knowledge of the meaning of whole sentences and that it is the task of the theorist of meaning to articulate a structure for that knowledge and to abstract stable meanings for the words that recur in a number of sentences. Speaker's knowledge of the meaning of words would be a theoretical abstraction from his knowledge of the meaning of whole sentences. These would be represented in the meaning theory by particular axioms.

This reading of Dummett concurs with what he says about an axiom earning its place in a theory only to the extent that it can be used for the derivation of theorems for whole sentences. It is the theorems which correspond to specific linguistic abilities by which speakers manifest their knowledge, but "there need not be any direct correlation of that knowledge which is taken as constituting understanding of any one word with any specific linguistic ability"; and so:

The ascription of a grasp of the axioms governing words is a means of representing his derivation of component words, but his knowledge of the axioms need not be manifested in anything but the employment of the sentences. (Dummett 1976, p71)

Does this mean that our understanding of each basic sentence in the language is initially unstructured, and that their inclusion of component word senses can only be appreciated once we make reflective judgements

Chapter Six

about how their component words are exercised in other, perhaps more complex, sentences? This seems unsatisfactory for several reasons. Firstly, because it turns the account of meaning understood on the basis of unmediated recognitional capacities into a kind of sentential atomism. But this conception of what is understood is in tension with the theorist's account of meaning; for while *speakers* grasp these meanings as a whole, the *theory* describes them as having parts. Secondly, there is simply not enough structure in those meanings to show how the speaker could be expressing a conceptual judgement about what he recognises. For without the thought that he could make similar judgements on other occasions, re-identify individual particulars in other scenes, and recognise other objects as possessing the properties ascribed to the objects recognised on one occasion, we should have reason to doubt that an exercise of a recognitional capacity involves conceptual thought at all. What would be the difference between a speaker of a feature-placing language recognising similar situations and speakers recognising the recurrence of an individual particular unless that recognitional capacity embodied a criterion of identification for that object? The point is that to have conceptual content, a recognition-based judgement has to be seen as the exercise of distinct and re-usable concepts; and to know which content a judgement involves we need to be told more precisely which concepts are involved. For this we need some account of the re-combinable parts of these recognitional judgements. Thirdly, not only should the details of our grasp of the meanings of basic sentences tell us something about the content and the structure of our ways of conceptually dividing reality into individual objects and properties, they also tell us something about the semantical organisation of a sentence. This is vital if Dummett's appeal to the faculty of unmediated recognition is to avoid the crudity of the positivist picture which simply associates each sentence with some set of verifying experiences.

All of these points suggest that to give an account of *what* is grasped Dummett needs to give an account of *the structure of what is grasped*; and for this his instrumentalist account of our knowledge of the meanings of words is hopeless. However, there is a second reading of Dummett where he seems to endorse the requirement of the antecedent, and where he does

Chapter Six

want to see the notion of a sentence's internal structure as entering into a speaker's conception of what is said by the sentence. This reading is established in countless places in his writings. (see Dummett 1975a, p111-2, Dummett 1976, p69, Dummett 1973, p229, Dummett 1981a, pp276-281, 310-2, 460-1, 1982, p105). Consider the following statement on the need for structured understanding:

...there is no such thing as understanding a sentence as a whole, without reference to its structure, since understanding a sentence requires an understanding of at least part of the language to which it belongs, and the understanding of language cannot be described without reference to structure. (Dummett 1981a, p310)

Even on the output side there is a need to recognise distinct components in understanding:

When...understanding is held to reside in a grasp of the truth condition of a sentence, this will be manifested by the ability to identifythe bearer of the name and to judge, at least of such objects, whether they [fall under a given concept] and the preparedness to accept that object's falling under that concept as determining the sentence as true. (Ibid., p311)

This places a heavy burden on what can be manifest in the exercise of an unmediated recognitional capacity. However, perhaps nothing like this is involved for the basic sentences Dummett has in mind. But even in the simplest cases of understanding he seems prepared to admit that something like this is required:

...recognition of the truth-condition for "Watling Street is straight" as being fulfilled should involve two elements, the identification of the object and the recognition that it satisfies the predicate. (Dummett 1981a, p312)

The two elements would seem to be the conceptual constituents of the recognition, and so to appreciate the conceptual content of the recognition in full we need some account of the structure and components of a recognitional ability. But as we saw above Dummett thinks that the meaning theorist can say nothing further about those recognitional

capacities beyond attributing them to the speaker. Of course the meaning theory can also describe the structure of a sentence whose meaning is grasped by means of a recognitional capacity, but to assert that the theory provides the structure of a speaker's understanding is to give up the pretense of describing the actual facts of a speaker's understanding and be merely attributionist about the structure of competence. To explain what a speaker's grasp of the meaning and structure of those sentences consists in will take us beyond the theory of meaning, for it must provide a non-truth-theoretic account of the structure of our recognitional abilities. Only then can the facts of understanding confirm the form of the theory of meaning. The account I recommend would see our recognitional capacities as psychologically structured abilities and we would then need to call upon the cognitive psychological notion of a causally explanatory structure in the speaker. But just note that in a last ditch effort to avoid any such non-linguistic explanation of structure, and despite his disclaimer that neither the meaning theorist nor the speaker can say any more about the recognitional capacities for basic statements, Dummett is prepared to say:

....knowledge of language is normally accompanied by explicit, though partial, knowledge of the articulation of the complex capacity corresponding to the articulation of sentences. (Dummett 1981a, p310)

Even though this may be correct, we are entitled to some account of the speaker's possession of that knowledge; and this must be a non-verbal account on pain of regress.

Dummett's difficulties arise, I suggest, from an attempt to avoid empirical psychological matters in the theory of meaning, and to give a language-first account of our grasp of concepts. Dummett is not unaware of the need to give accounts of our grasp of particular concepts like red, or square, and of our knowledge of particular things; and he is surely right to tie these accounts to our possession of certain recognitional abilities. But he is unwilling to give a non-linguistic account of these concepts and our grasp of them, and it is this which leads him into regress. For the only account that Dummett can give of a speaker's ability to recognise red

Chapter Six

things, or his recognitional ability for a particular person X, consists in the speaker's being able to judge, when in a deciding position, the truth of recognition statements like 'This is red' and 'This is X'. But once again, the question arises about the nature of our grasp of the meanings those statements. Is it structured? Do we have to grasp the meanings of the constituent expressions? In particular, what is Dummett's model of a speaker's grasp of the demonstrative expression in the subject position of these statements? As Evans points out, it had better not take the same linguistic form as before, on pain of vicious regress (Evans 1982, p95). It is because Dummett wishes to account for our grasp of the the meanings of words, and the concepts they express, in the linguistic context of our use of whole sentences including them, that he is reluctant to say anything more about the meanings of those sentences or the component conceptual abilities exercised in our understanding. In trying to explain the meanings of words via the meanings of sentences, without begging the question as to what constitutes speakers attaching the meanings they do to their sentences, Dummett has to explain what it is for speakers to grasp the meaning of whole statements without drawing on the details of their parts. Like Frege's contextual definitions his problem is to explain what it is to attach a particular sense to a whole proposition. But as Dummett presents it, the account is either incomplete, or inconsistent with much else that he says about sentence understanding.

How is he to characterise sentence understanding at first without reference to sentence parts? One way forward here is behaviourism. We could take a speaker's understanding of a sentence S to consist simply in his dispositions to suit his use of it to the patterns of assent and dissent shown in the responses of others. Thus 'X understands S' is true provided that X would act in accordance with S's meaning. When we combine this with a theory of truth which serves as a theory of meaning, we get 'X understands S as true when and only when p' is true provided that X would act in accordance with the conditions for S's truth; ie. would assent to S when and only when p. But this is hardly Dummettian antirealism which demands that a speaker's understanding should consist in his exercise of *cognitive* capacities. Thus Dummett would say for some basic sentence S, that 'X understands S as true when and only when p' is true

Chapter Six

provided that X is able to recognise the conditions p for asserting S. But now what account of that recognitional capacity can distinguish it from behaviourism. Moreover, what account can show that it confers understanding which can also be described as the grasp of something structured, and, where relevant, as involving grasp of some expression in S whose use is indexically tied to some particular part of the recognised scene? Without prior appeal to the subject's grasp of the concepts and structures involved, Dummett seems to lack the materials to offer any satisfactory account here.

Certainly, at times Dummett points to the non-behaviourist elements of his approach, when he acknowledges that at the primitive level of language use (that which includes demonstratives and indexicals as the principal devices for making reference) we are operating with perceptual procedures which "involve an interplay of sensation and conceptual thought" (Dummett 1981a, p144) But he goes no further towards a non-linguistic account of our handling of indexicals, and he offers no non-linguistic account of the structure of conceptual thought. So we are given nothing more to illuminate the character of linguistic grasp. Without such materials he cannot give a convincing account of the distinction he needs between what we might call *weak knowledge* and *strong knowledge* of truth conditions. Weak knowledge of truth conditions is possessed by a person or a device which can detect and register the obtaining of conditions for a sentence's truth. This amounts to merely knowing when a sentence is true. In contrast, someone has strong knowledge of a sentence's truth conditions when he grasps the structure of the sentence and its constituent senses. This knowledge requires a speaker to know what the sentence is about, in the sense of knowing which objects and properties its terms and predicates are about (Russell's Principle). Now whereas weak knowledge of truth conditions requires only that the speaker or device suits its dispositions or practice to the output theorems of a T-theory, strong knowledge of truth conditions requires a speaker to know what it is for the sentence to have that truth condition. For the attributionist, this might be knowledge of the theory from which the theorem is derived, but for Dummett it must be something closer to the speaker's understanding of the sentence itself; that is, knowledge

Chapter Six

embodied in the specific linguistic ability correlated with the relevant theorem. However we characterise that ability, to count as possession of strong knowledge of truth conditions it must confer on its possessor an understanding of the sentence that involves "grasp of how it is determined as true, if true, in accordance with its meaning" (Dummett 1982, p61). It is in the attempt to account for these details of grasp that the difference between Dummett and the interpreter shows. For unlike the interpreter's externally attributed semantical descriptions, Dummett thinks that:

a theory of meaning must ascribe to a speaker an implicit knowledge of the underlying semantic theory (Ibid. p62)

But since that knowledge can only be meaningfully ascribed when it is manifested in exercise of specific linguistic abilities - abilities correlated with output theorems of the semantic theory - we cannot simply use the theory's description of those abilities to confirm that the speaker has strong knowledge of (structured) truth conditions. Dummett needs some non-attributionist way of meeting Wright's Challenge. An account of the psychological facts about grasp and the conceptual abilities exercised in knowledge of any particular object or property is what is needed here. Without these it is hard to see how we can do better than account for speaker's weak knowledge of truth conditions. To settle for an account in terms of weak knowledge does not really tell us what *speakers* grasp since we have to rely on the interior of the theory to put any gloss on the sentence on the theorem's right-hand side. To serve as a theory of understanding we need more substantive conditions on what it is for the speaker to grasp the truth conditions the meaning theorist assigns to his sentences. The trouble is that by resisting the theory of thought and without an account of cognition Dummett lacks the materials necessary to complete an account of our understanding of a structured language and our possession of concepts.

Ironically, the only other alternative to the Davidsonian picture of attributed understanding comes from John McDowell's view of speakers' understanding as a perceptual capacity. Although McDowell modestly presupposes the contents of the language he describes, by adverting to this

Chapter Six

perceptual capacity he is attempting to offer us a thesis about the character of a speaker's *own* knowledge of meaning. For McDowell:

Understanding a language consists in the ability to know, when speakers produce utterances in it, what propositional acts with what contents they are performing. (McDowell 1976, p45)

This ability is quite simply an ability to hear utterances as speech acts of a certain kind with specific contents. This is non-inferential knowledge with the distinctive phenomenology of the unreflective language user. But why then should McDowell also be interested in giving a formal theory of truth to describe speaker's competence? It would seem rather roundabout as a description of their knowledge if they do not actually make use of it. McDowell's reply would be that we need to characterise the perceptual capacity to know precisely what it is a capacity to do, or rather which capacity it is; and for an interpretation theorist like McDowell, the truth theory characterises the relation that defines the capacity. But where in all this do we earn the right to describe speaker's sentences as semantically structured in the way the T-theory describes; and how does McDowell guarantee reference to sun-sentential expressions? The answer is that McDowell is also an attributionist about speaker's knowledge. The content and character of the theory of meaning give the content and character of a speaker's knowledge of his language. It is the theory which gives a conceptual character and structure to the speaker's perceptual capacity. While it is just plausible to be an attributionist about the articulation of a speaker's abilities and hence his knowledge of language, reading-off the the separate pieces of a speaker's knowledge from the clauses and theorems of the meaning-theory, it is less clear how McDowell can substantiate his claims for the reference of words and their individual senses. If truth-theories are anchored to the evidence at the level of their theorems not their axioms, then being an attributionist about the structure of a speaker's understanding of sentences will guarantee separable components within that knowledge, but this move alone cannot substantiate claims for reference. But no further account is offered to guarantee genuine empirical import to the relation of reference. For a guarantee of properties of reference for a speaker's words we must rely on

the modesty of McDowell's theory of meaning. But this means no account is offered at all, just an unsubstantiated presupposition that the words do make reference to extra-linguistic items. What is so unsatisfactory about this position is that it smuggles in an intended interpretation (in terms of realist truth conditions) for the words and sentences that appear with what is supposedly their normal uses on the right-hand sides of the axioms and T-sentences. But this is merely question-begging, in that the antirealist could in similar vein simply assert *his* preferred reading for the right-hand sides, in terms of the notion of assertibility. Neither realist nor antirealist would have given a non-question begging argument for his conclusion. The realist/antirealist debate should not be begged in this way, and certainly not by appeal to one's favourite theory of meaning, given that the whole question of realism and antirealism is an issue of debate in the theory of meaning.

It is now time to motivate an explanatory approach to these matters which will substantiate claims for knowledge of meaning and structure without begging questions of content and form. This would alleviate the needs of the Descriptivist in the unstable middle position who is trying to resist attributionism by looking for facts about speakers' knowledge for the theory of meaning to describe. For this we need an explanatory theory of those mental states of a speaker - his knowledge of language - that enter into a causal explanation of his language-mastery.

Chapter Seven

The Indispensability of Concepts

7.1) The System of Concepts

We saw above how the interpretation theory leads us to the surprising conclusion that talk about concepts is purely instrumental. Davidson is a realist about attitudes - there are facts of the matter available to interpreters (up to indeterminacy of schemes) about when it is correct to ascribe a set of attitudes to a subject - but he is an instrumentalist about concepts as constituents of their contents.

But is this even coherent? Some think not. Gareth Evans, for one, thinks that belief involves the possession of concepts:

Behind the idea of a system of beliefs lies that of a system of concepts, the structure in which determines the inferential properties which thoughts [possess]. (Evans 1981, p132)

But is the inference from a holistic web of inferentially connected thoughts to conceptual structuring in those thoughts too swift? Let us start with the premiss. The ascription of even a single attitude imposes a number of requirements on a thinker. For Davidson, one belief makes sense only in the presence of many others to which it is naturally tied, and to whose presence it owes its own particular content. The requirement that other attitudes be actually present in a thinker may be too strong, but Evans, who does not think of concepts as elements of thoughts, but rather as conceptual capacities or abilities we exercise in thought and action, sees the ability to think certain thoughts as requiring that the thinker be capable of thinking certain others. And he explains this requirement by means of the underlying system of concepts because he believes that thinking thoughts amounts to the joint exercise of a complex of interacting abilities. Arguing in this way he thinks can discern a general constraint upon what it is to be capable of thinking a thought with a propositional content (or being ascribed a propositional attitude, or understanding and assenting to a proposition that expresses that thought.) He calls this condition the Generality Constraint:

Chapter Seven

If a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception.We thus see the thought that a is F as lying at the intersection of two series of thoughts: on the one hand, the series of thoughts that a is F, that b is F, that c is F,...,and on the other hand, the series of thoughts that a is F, that a is G, that a is H,..... (Evans 1982, p104 & fn.21)

From this Evans draws the conclusion that being able to grasp the thought that a is F involves the exercise of two distinct abilities; for example, it involves "knowledge of what it is for something to be F - which can be exercised in indefinitely many distinct thoughts, and would be exercised in, for instance, the thought that b is F." (Ibid., 103) There must also be an ability to think of a particular object, for this is "a capacity which, when combined with a knowledge of what it is in general for an object to be F, yields the ability to entertain the thought that a is F, or at least a knowledge of what it...would be for a to be F" (Loc. cit.) These knowledge-involving abilities are exercised jointly and only in the context of thinking whole thoughts, but each one is exercised in a number of different thoughts in tandem with yet other (though categorially appropriate) abilities, and it plays a common role in the explanations of our grasp of these different thoughts. Thus thoughts are structured, and by this, Evans means, conceptually structured, and not merely supported by cognitive mechanisms that are causally structured - although this could be true too - for the component abilities that enter the explanations are knowledge-involving.

If it is sound, this is a powerful argument for the conceptual structure of thought, sadly lacking in the Davidsonian story. The best Davidson can do is try to interpret people's speech as expressing their beliefs, and other thoughts, and then try to induce semantic structure on their sentences through the use of an interpretive truth theory for their language. The structure of language could then be treated as the best guide to the structure of thought. But we have seen how insubstantial Davidson's claims are for linguistic structure. The strategy of interpretation is simply not committed to thought being essentially structured. But Evans is.

Chapter Seven

As we saw above Dummett is a language first theorist and appeals to the Fregean assumption to substantiate and explain the structure of thought. This commits Dummett to having to substantiate linguistic structure (but unlike Frege in a non-platonistic way). Evans cannot follow this strategy because he is a thought-first theorist and as a theorist of thoughts he separates the structure of thought from the structure of language. He thinks that "while sentences need not be structured, thoughts are *essentially* structured" (Ibid, p102). He is thinking of elliptical sentences, eg. "Fire", where a complex meaning is expressed quite simply. But the strong conclusion he wants is unmistakable:

it is simply not a possibility for for the thought that a is F to be unstructured - that is, not to be the exercise of two distinct abilities. (loc.cit.)

If he is right about this then Davidson's interpretation theory falls down:

We should surely be reluctant to assign the content 'I am in pain' to any internal state of a subject unless we were persuaded that the subject possessed an idea of what it is for someone - not necessarily himself - to be in pain, and unless we were persuaded that the internal state in question involved the exercise of this idea. (Ibid, p103, underlining mine)

But is Evans entitled to these conclusions from the Generality Constraint? Can we argue from this to the structure of thought? Martin Davies's careful work has shown us that we can't. As Davies points out, the Generality Constraint only commits us to a certain closure condition on the contents of thoughts, which he illustrates as follows:

if a thinker can think that a is F and that b is G, then the thinker is able to frame the thought that a is G and that b is F; and if a thinker is able to think the thought that a is R to b, then the thinker is able to frame the thought that b is R to a. (Davies 1989 p15)

Chapter Seven

This thought allows us to see how the conclusion of a transitive inference $R(a\ c)$, from $R(a\ b)$ and $R(b\ c)$ is made conceptually available to a thinker in virtue of his grasping the premisses. (Why he chooses that conclusion requires a separate explanation.)

It seems to me that Davidson could accept this closure condition, and that interpretation theory could satisfy it without yet committing Davidson to the essential structure of thoughts. The important question to ask at this stage is why Evans thinks we should be committed to it. It has been accepted that the Generality Constraint is a necessary condition on thought, and it might even be accepted that Evans is offering an explanatorily insightful way of satisfying it in terms of these complexes of component abilities. But the Davidsonian will want to know why this is obligatory. Davidsonians would also point out that the explanatory story is irrelevant to the conceptual issue of whether or not we have concepts. (No pun intended). They will say that this is an implementational matter. How we satisfy the Generality Constraint may be explained computationally but that we satisfy it is all that matters.

It is important to tread carefully here because Evans is not saying that the structure involved here is just a computational matter, even if it can be illuminatingly explained by a computational story.

Once again, Evans, like McDowell, could warn of a kind of 'naturalistic fallacy' in the offing, here, if anyone thought the structure of a computational processor that conformed to the Generality Constraint was what it was to satisfy that constraint. Plenty of computational devices could exhibit the sort of closure property such Generality involves, all that is needed is a device that can point to, or move through, any of the interstices of a Cartesian coordinate framework. This is not difficult to explain. What it is conceptually difficult to explain, without a reduction, is why a certain structured process amounts to the exercise of *knowledge-involving* abilities. Evans is not attempting a reduction and he is still working at the level of knowledge. So what is behind Evans' insistence on the essential structure of thought? It might be that Evans treats *conceptual structure* as a necessary condition of thinking:

It is a feature of the thought-content *that John is Happy* that to grasp it requires distinguishable skills. In particular, it

Chapter Seven

requires possession of the concept of happiness - knowledge of what it is for a person to be happy; and that is not something tied to this or that particular question. (Ibid., p103)

If Evans just means to make this a constitutive claim on what it is to possess certain thoughts, and the Davidsonian refuses it as such, we seem to reach an impasse where they simply trade rival constitutive claims. This makes for very unproductive and sterile Oxford Philosophy. To avoid this I shall suggest a controversial way out.

What is not controversial is that Evans embraces what I shall call the Cognitivist Assumption, believing that the theory of our cognitive underpinnings enters into the top-level theory of our knowledge or grasp. This is because the notions the highest level deals with are to be located at the level of our cognitive mechanisms and because the cognitive theory shows us how the requirements on these notions are met. (This is not to say that the cognitive theory of our internal mechanisms can stand alone: it can't. In fact, adoption of the Cognitivist Assumption means that one thinks that the explanatory space occupied by such a meagre theory should be occupied by a conceptually richer explanatory theory of content that draws upon a good deal of the details of a theory of mechanism. A slightly different way of making the same point is to say that a meld of the conceptual content theory and the cognitive processing theory, originally conceived as theories at Marr levels 1 and 2, respectively, is to be achieved by moving to level 1.5. This is the tack taken by Peacocke in adopting the cognitivist position. (Peacocke 1986b).

7.2 Causal Systematicity

I think that it is in this cognitivist spirit that Davies provides Evans with an argument to the structure of thought from what he calls 'the causal systematicity of inference'. This may go beyond anything Evans explicitly claims but there are hints of such a leaning in the Varieties of Reference and "Semantics Theory and Tacit Knowledge" (Evans 1981).

Davies' idea goes like this. Being a believer involves the capacity to make systematic inferential transitions between thoughts. This is just the closure condition and even Davidson would agree with this. But are the

Chapter Seven

processes that realise these transitions causally systematic in a way that corresponds to the systematicity of our inferential capacities? Davies thinks that if they are then the structure of processing offers us good reason to think that there is structure in the inputs and outputs of that process. So making inferences would require the thinker to have structured thoughts.

The reasoning is concerned with the conditional claim that if a cognitive process is systematic then the inputs and outputs to that process are syntactically structured; ie. there is an inference from the system to its syntax. Although it is a conditional claim, to show this is to establish an important result because it argues for a conclusion about our cognition from an empirical claim about our cognitive architecture. Here is Davies' argument. He starts with the antecedent.

Processes mediate between inputs and outputs. Processes are systematic relative to patterns in their input-output relations. In a cognitivist/computational setting, this pattern is revealed under a semantic description of the inputs and outputs to the process. Now suppose that we are dealing with aspects of human inference making and suppose we notice that there is a common pattern in the inferences people make from the thoughts 'Margaret Thatcher alienated her European partners', and 'Margaret Thatcher lost support at home', to 'Margaret Thatcher has both alienated her European partners and lost support at home', and from 'Tessa Blackstone is the Baroness of Stoke Newington' and 'Tessa Blackstone is the Master of Birkbeck College' to 'Tessa Blackstone is both Baroness of Stoke Newington and Master of Birkbeck College'. Several input-output pairs may exhibit this common pattern; so we can frame a semantic generalisation which we can state as "Whenever there are thoughts of the form *a* is *F* and *a* is *G* one can infer a thought of the form "*a* is *G* and *F*." The semantic generalisation goes beyond mere conjunction reduction because some transitions are illicit; eg. those from "Someone is in love with him" and "Someone has his welfare in mind" to "Someone is both in love with him and has his welfare in mind". The semantic generalisations will identify a pattern of conjunction reduction inferences where it is the same person being referred to in the two conjuncts.

Chapter Seven

The cognitivist/computationalist will ask the question, do these various input-output transitions have a common causal explanation corresponding to the pattern they instantiate? To deny a common causal explanation is to think of each of these transitions from pairs of thought-conjuncts to the conjunctively reduced thoughts as being separate dedicated processes; hard wired, if you like.

Whereas if there is a common causal explanation in all cases of inferences from thoughts described as having the semantic form: a is F, a is G; to thoughts described as having the semantic form: a is F and G, then we may say that the process of inference is *causally systematic relative to that pattern*, ie the pattern captured by our generalisation. Just conforming to that pattern in inputs and outputs is not sufficient for causal systematicity: two machines could exhibit the same input output patterns but still be causally systematic with respect to different patterns; calculators are the obvious examples. So we cannot conclude that systematic processors with a certain input-output patterns are causally systematic with respect to that pattern.

As Davies puts it:

A systematic process is not merely a process whose input-output relation conforms to such patterns as these. Systematicity relative to these patterns requires that, corresponding to each pattern (in extension) there should be a real commonality of process: a causal common factor in the form of a component mechanism implicated in the various transitions that conform to the pattern. (Davies forthcoming, p7)

That a device can be given a systematic semantical description does not entail the systematicity of its internal processes. So in our case, there is no direct argument from systematic descriptions of the semantic contents of our mental states - our having thoughts which meet the Generality Constraint - to our having structured processes that mediate the transitions between them.

In accordance with the Davidsonian story, inferences can be made from propositional attitudes to actions. Consider an example of this where there is a choice of outputs. Suppose that Sheila has the enduring desire to avoid seeing her husband Bill. Unfortunately they work together in the

same office. Now suppose that she is sitting in a restaurant at lunchtime deciding what to do. It makes sense to suppose that in concert with her deeply held desire to avoid Bill she can have various beliefs, each of which give her a different reason to act. Imagine the following four cases. (1) Sheila believes that Bill will take a bottle of brandy back to the office, so she orders another coffee and a taxi to take her home. (2) Sheila believes that Bill will take a bottle of brandy home, so she orders another coffee and a taxi to take her to the office. (3) Sheila believes that Bill will want several more brandies before turning up at the office plastered, so she refuses another coffee and orders a taxi to take her home. (4) Sheila believes that Bill will want several more brandies before turning up at home plastered, so she refuses another coffee and orders a taxi to take her to the office.

Davidson could rationalise this behaviour in intentional terms: these are the patterns of a disastrous relationship. But as far as the processing story goes, there are (at least) two very different ways to describe the intentional causes of her action. Let us also suppose that her desire for a second cup of coffee is constant too, but her desire to be rid of Bill is even greater. The two possible descriptions are as follows: either each complex of belief and desire just leads to its unique outcome in independence of the others (she has a set of dedicated intentions); or there is a causally efficacious common factor in the explanations of her refusing a second cup of coffee, and a different causal efficacious common factor in explanations of her having a second cup. (We can also add to the latter sorts of explanation, the cross-cutting causal common factor in the explanations of cases where she goes to the office, or the causal common factor in explanations where she goes home.) The cognitive psychologist would want to ask whether the causally efficacious reasons for Sheila's acting as she does are causally *systematic* or not? Is there something in common to the explanations where she has to refuse a second cup of coffee? Certainly in the intentional descriptions of the inputs they are both cases where part of her belief has to do with the likelihood of *Bill's having a brandy*. And in the cases where she goes to the office, what she does has to do with a part of her belief concerning the likelihood of *Bill' not going there*. But these patterns are due to semantical descriptions of the input-output pairs.

Chapter Seven

What about the states of belief themselves. We need to know whether there is a common explanation for her actions when her beliefs concern Bill's having several more brandies, or a common explanation when her beliefs concern his destination after leaving the restaurant. And, in cases such as these Davies says: "The answers to these questions are not determined by the facts about the input-output relation, but by facts about the internal [cognitive] architecture of the [agent]" (Ibid, p6 brackets mine.)

Although we now know what is is for a cognitive process to be causally systematic, why does causal systematicity entail syntactic structuring in the inputs? First we need to know what syntactic structuring at this level amounts to.

In a computational setting syntax is Janus-faced. Looking outwards, what counts as a syntactic property of the input depends on what semantic properties we can discern; syntactic properties are systematically related to semantics. Looking inwards, the syntactic properties of an input state depend on the internal constitution of the machine; the causal role of a symbol string in one machine may be quite different when it is input to another. For example, what counts as an assertion in one programming environment, may count as a procedure call in another. (Students learning to formalise English sentences in propositional calculus pay scant attention to their internal structures; later when they have to formalise in predicate calculus they are forced to change. Once again a syntactic property is relative to a programming environment. Syntactic properties are doubly relative.

If we suppose there is causal systematicity in our example we can see that causal systematicity imposes requirements on causal properties of the input states. Each time Sheila orders a taxi for home, there must be some causal component that makes her act in that way. And each time she has a second cup of coffee there must be something that makes her do that too. In the cases where she orders a taxi to take her to the office we can suppose that it is the causal property common to the belief state with the content that Bill will take a bottle of brandy home, and the belief state with the content that Bill will have several more brandies before going home. But whatever this is, the two belief states do not have all their causal powers in common: there is a causally relevant difference between them because

Chapter Seven

in one case she orders another coffee and in the other, she refuses a second cup. So there is one commonality and one relevant difference in the causally explanatory powers of the two belief states.

As Davies points out, it is an empirical question what these causally salient properties are, but if we suppose that this happens to Sheila quite often, it is hard to believe that the commonality is just a physical similarity each time. What we should say is that on every occasion of a certain intentionally described type, where there is a constant complex of beliefs, desires and actions, there is a physical property that causes her to behave in these intentionally similar ways, but not the same physical property each time. So what are the constraints on the physical properties that can achieve the right result? Well, they have physically to realise the causal role of a state whose operative properties produces certain input-output patterns in the agent as described in intentional terms. We can now describe the following set-up. Causal systematicity in a decision-making agent requires the belief states that are input to a process of reasoning (the same argument could run for her desires) to have causally operative properties which correlate with semantical descriptions of those inputs. So the belief with the content that Bill will have several more brandies (before going home or to the office) should have something in common with other input belief states that engage whatever mechanism makes her say no to the offer of coffee and leave the restaurant in a taxi. Over all the inputs and permutations of cases, these correlations between semantical properties and causal properties guarantees that causal systematicity keeps in step with causal systematicity. But what property corresponds to an aspect of the semantic contents of beliefs that is constant across all input states that cause her to leave so quickly? Whatever property it is it must also affect her internal operations if it causes her to act like that. The property we want is the doubly relative notion of a syntactic property. This establishes the conditional claim. Davies point out, that it requires neither that every aspect of semantic content have a corresponding syntactic property, nor that the inferential transitions should be explicitly represented.

This is just a conditional claim so now we want to know if we can detach. Is there any reason to suppose that some of our cognitive

Chapter Seven

processes are so causally systematic? Let us return to Evans and his claim that underlying our system of beliefs there is a system of concepts. If this were true we could put more of a gloss on the Generality Constraint and the closure condition it entails. We could say that the domain of conceptual thinking had to be closed under recombination of *syntactically appropriate* conceptual constituents. To be capable of thinking the thought that a is F, one must have the resources for thinking the thought that b is F, that c is F, and so on, for all particular objects of which we have a conception. Having mastery of the concept F, in general, is knowing what it is for an arbitrary individual to be F, in this way we do not need to account for a subject's capacity to know what it is for his thoughts about F to be true one by one. As Evans says, knowledge of what it is for something to be an F "can be exercised in indefinitely many distinct thoughts, and would be exercised in, for instance the thought that b if F" (1982, p103). He also says (Evans 1981) that to have a belief requires one to understand its location in a network of other beliefs, and this depends on the thinker grasping certain inferential links:

the subject's appreciation of the inferential potential of one belief (eg. the belief that a is F) at least partly depending upon the same general capacity as his appreciation of the inferential potential of others (eg. the belief that b is F) (1981, p132.)

If the very same capacity was exercised in making transitions from say a is F to a is H and in transitions from b is F to b is G, Evans would say that possession of that capacity would amount to mastery of a concept. We can put this by saying there is a causal systematicity:

the two inferences are manifestations of a common underlying capacity: namely mastery of the concept of being F. (Davies 1989, p15)

Another way to describe this systematicity is to point out that to think the thought that something is F one must grasp the concept of F, ie. know what it is for something to be an F. In particular this may involve knowing that all Fs are Hs, and therefore there is an inferential transition we are prepared to make on thinking or discovering that something is an

Chapter Seven

F, namely that it is an H. The causal systematicity explains this common inferential capacity of F-involving thoughts, because it provides a mechanism for the transitions, and so it can play the major part in explaining what it is for us to embody grasp of the concept F. The mechanism can be a crude causal mechanism, so the device that effects the transitions need not be content-involving but they have to operate in virtue of properties of the input that triggers them, and this now argues for a syntax systematically related to the conceptual structure of thought.

This is Christopher Peacocke's position (Peacocke 1989u) where he says that if we reflect on these inferential transitions we will admit to finding them primitively compelling; and what is more they will be *primitively compelling in virtue of their form*. Inferential transitions would be part of what Peacocke calls the possession conditions of a concepts. Possession conditions offer us a way to give a more direct and less holistic characterisation of what it is to have mastery of a given concept. Together with the arguments in Evans and Davies, this is a way of responding to the attributionist's claim that there is no *other* way to ground the ascription to an individual of a particular concept than to say that it pulls its theoretical weight in making attributions of beliefs and desires that can be confirmed by holistic interpretation.

These can all be regarded as explanatory approaches to concept mastery in that they seek to give a direct account of what it is to possess a concept, or to put constraints on the form any such account must take. These are adequacy conditions of theories of thought. In offering positive proposals on the nature of concept possession they certainly offer an alternative to the the attributionist's picture of things. But do they go beyond the claim to offer a competing rival account? Yes they do. I think this is clearest in the case of Davies who is arguing for an internal architectural constraint on the contents ascribable to a system. These amount to internal architectural constraints on what it is to be a believer: to lack this inner cognitive organisation is to fail to satisfy the requirements on what it is to be capable of conceptual thinking and reasoning.

This provides us with the desired contrast with Davidson's attributionist minimalism. Davidson says that:

Chapter Seven

the question how to tell when a creature has propositional attitudes.....is not empirical; the question is what sort of empirical evidence is relevant to deciding when a creature has propositional attitudes. (Davidson 1982, p317)

As we have seen, he concludes that it must be behavioural evidence, including evidence of linguistic behaviour. Davies, Evans and Peacocke are all arguing for cognitive conditions of internal structuring on what it is to be a thinker, and not merely behavioural constraints.

The difference can be brought out like this: whereas Davidson would say that if two creatures shared all the same behavioural dispositions and physical histories, they would have exactly the same mental lives, Davies is committed to the idea that they could share all the same input-output patterns in their environments and behaviour, and yet be causally systematic with respect to different semantical or intentional patterns in their input-output pairs. In such a case Davies would say that they had different mental lives. For Davidson, the mental supervenes on the behavioural, physically conceived. For Davies *et al*, the mental life of human beings supervenes on their behaviour and their internal cognitive structuring.

I find Davies' argument convincing, with the caveat that the syntax in conceptual structure is rather more complex than these examples suggest. Davies is aware of this but I suspect he would draw different lessons from this than I would. For in contradistinction to Evans, Peacocke, and perhaps Davies, I do not think that it is correct to attempt to base a theory of meaning for a natural language on a prior theory of thought. This is Evans strategy. He hopes to explain our ability to understand various linguistic constructions by reference to the kinds of thoughts we have to think to understand them; and for this he needs a prior account of thoughts and what our grasp of those thoughts consists in that does not refer him straight back to our ability to use various natural language sentences. To assert that we do not need and possibly cannot have a theory of thought prior to a theory of the language in which that thought is expressed, is not to argue for an explanatory priority that forbids all access to the level of thought in our theorising about language. It is just to assert that we could not have a prior and finished theory of thought in whose terms we could

give an account of public language meaning. Thought cannot be exhaustively explained without reference to explanations of linguistic significance. This is not primarily for reasons of the publicity of thought and meaning, although some ranges of thought cannot be expressed or are simply unavailable without prior mastery of the use of linguistic expressions. Rather, they have to do with the fact that the syntactic structure of conceptual thought must be at least as intricate as the syntactic structure in language, which requires symbol strings to have not just constituent structure but also hierarchical structure-dependencies. Those constituents that serve as basic categories must also be understood in terms of their syntactic roles in larger structures; all of which suggests a theory of grammar at least as complex as that of natural language syntax. My argument would be that syntactic structure of this complexity exists because of the systematic role it plays in relation to the semantics of languages.

Chomsky has argued that the primary role of language is not communication but expression, and syntactic structuring is certainly an important ingredient in facilitating the huge range of expressions available in natural language. Chomsky has also argued that the basic principles of this syntactic structure are innately possessed, encoded in certain cognitive mechanisms dedicated to the production and comprehension of language. Those thought theorists who thought the language-like propositions that inhabit the domain of conceptual thinking were available before our capacity to speak a language would be involved in a merely terminological dispute with me. The very capacity for thought they described would need to make an essential appeal to part of our innate equipment for mastering a public language. It would be truly a language of thought!

Evans acknowledges the point about the categorial constraints on the recombinations of conceptual constituents, but does not pay sufficient heed to the point when dismissing an account of thought-constituents in terms of elements in a structure, in favour of structured capacities.

I think one of Evans' great achievements was to point the range of non-conceptual contents in our thinking and to trace out the connections between these and certain informational underpinnings and the contents

Chapter Seven

of conceptual thought. Importantly, he draws attention to how certain thought-contents are partly determined by their connections with these others notions, and just how big a part certain informational connections with the world play in the possibilities of reference and singular thought.

Unlike those who suppose a strict explanatory priority of language over thought, I would suggest an ecumenical stance here. There may be many places where we have recourse to notions that find their home in the theory of thought and theories of the cognitive psychological underpinnings to language-mastery, and I suggest that no adequate theory of linguistic understanding can be given without drawing upon them. But this is not to recommend a thought-first approach to the problem of meaning.

There may be many kinds of psychological contents and levels of structure which support and sustain our ability to use and understand a natural language, but the meanings of items in public language cannot be reduced to any one of these; however, the possibility of meaning may require certain relations between them. One of the biggest tasks in the philosophy of language will be to construct a stratified account of linguistic content and to decide which psychological facts determine which aspects of meaning and which are due to facts of some other sort. I see no reason to think we can have a simplified unified account. The important point for the Davidsonian to grasp is that any such account will have need of a theory of concept mastery.

A good question is whether Evans can use his cognitivist credentials to undermine the Davidsonian position? Could he, for example, point out that a structured processing story is the only way to explain our satisfaction of the Generality Constraint and claim that a theory of thought (interpretative theory of the propositional attitudes) that could not meet the constraints of a theory of our grasp of thoughts could not be a plausible contender? Davidson should accept this challenge, for even if he did not share the cognitivist assumption that the theory of grasp had anything to do with the mental, he must concede that there is some non-mentalist story to be told about implementation. So, can there be brute processors whose workings explain how we can satisfy the conceptual constraint we must satisfy?

Chapter Seven

It is here that connectionism, or parallel distributed processing, in particular and Davidson's interpretation theory find themselves as strange bed-fellows. For the claim is that using non-symbolic representations, and hence non-symbolic processes, a connectionist network could be built to behave, in many respects as we do, while at the same time satisfying the closure condition. Different stable patterns of activation across the units in the system would serve as distributed representations of whole thoughts, and the systems could make transitions between these without there being any part of one of those patterns which corresponded on each occasion to a conceptual constituent that recurred as a common component, according to the structured account, in a each of set of other thoughts reached by this process. The process could be the superimposition of one pattern on another, where one patterns of activation can prompt another; an easy thing to achieve in connectionist computation. However, the key point would be that neither the structure of the representations or the processing need correspond to the Evansian structure of thought. There is no articulation of the states of the system into syntactically structured vehicles of representational content. This may provide us with the implementation of a believer who lacks concepts.

This is the claim anyway. But what still has to be established is whether connectionist networks are capable of satisfying the closure property, or producing the right outputs for the right linguistic inputs. Not just the input-output patterns of systematic inference matter here, but also patterns of breakdown and certain systematic failures. All this is part of satisfying the patterns in the right way, which is distinct from merely coinciding *de facto* on a number of occasions. Whether or not the connectionist networks are so systematic may turn out to be an empirical question. If the answer is no, Evans wins the day.

(Note also that this merely implementational role for combined connectionism and interpretation theory would not satisfy connectionists with genuinely *cognitive psychological* explanatory ambitions. An ambitious connectionism would want to explain the intentionality of mental life or the nature and content of our cognitive states. And in this

Chapter Seven

respect they would be rivals to Davidson because the two approaches would jostle for the same explanatory space.)

The hard task for the cognitivist's construal of linguistic structure is to say why the internal facts about a language that give structure to our capacity to understand our native language are genuinely psychological facts. For if the structure of language depends upon the structure of our knowledge of language, it must be shown that the structurally salient facts about linguistic processes are part of a speaker's *knowledge* of language. This has still to be shown.

Chapter Eight

Knowledge of Language

8.1) Knowledge of Language and Theories of Language

This thesis began by locating its concerns within a certain family of philosophical issues. These have to do with the relationship between language and mind. I said at the outset that all parties to the dispute that concerns me accept that there is an intimate connection between mind and meaning because of the mind's involvement in meaningful speech. Where they disagree is in their conceptions of the mental states involved in language mastery. I looked at one fully worked out view of the philosophy of mind and language called attributionism, and then went on to sketch the foundations for an alternative form of mentalism. And in turning now to theories of language, it is time for me to be more precise about the nature of this relationship between language and mind.

I suggest that we should locate theories of meaning of the sort we have already looked at within the larger project of constructing theories of language. Let me say something quite neutral to begin with about the point of a theory of language. Language, like meaning, is an elusive notion and is best approached via theory. But what does a theory of language look like? What are its data? What is such a theory for? Let me take these in order.

By the term 'theory of language' I shall mean a collection of complementary sub-theories, each of which addresses its own proprietary level of description of the linguistic data, together with a theory of the relation between the levels. So, for example, a theory of language will include theories of phonetics, phonology, morphology, syntax, semantics and possibly pragmatics. What are the linguistic data for such a theory?

The primary data for linguistic theories are examples of spontaneous speech. Speech is prior to writing in the order of development and more widespread than literacy so it constitutes our starting place. Pre-theoretically, we are not entitled to claim that a language exists independently of speech, except, of course, in derivative forms, like written inscription.

Chapter Eight

A secondary source of evidence for linguistic theories is equally available in behaviour but less commonly available than the first. This is a native speaker's intuitions about his or her own language. Described without mentalist overtones, linguistic intuitions amount to behavioural responses to specific linguistic promptings. Even Quine's behaviouristic field linguists rely on this. A whole range of data across subjects can be gleaned by posing questions that require speakers to produce 'acceptability' judgements on a range of specially contrived examples. It is of course a theoretical issue what this secondary source of native speakers' judgements of 'acceptability' are judgements of.

Thirdly, we can usually produce comparisons across populations of language users of the relative frequency of occurrence in their language of certain constructions, lexical items, phonemes, etc. These frequency and distribution data can also be achieved for single speakers if we consult a sufficiently large documented corpus of their speech.

Fourthly, we can look at cross-language comparisons, comparing the phonological or sound systems of those languages, the sizes of their vocabularies, the distributions of their terms within sentences; and then once we have constructed theories, of say, syntax, we can compare the syntactic constructions different languages permit by inventing examples of sentences which we test against the primary and secondary sources of evidence.

Fifthly, we have information about patterns of development from child language acquisition research.

And finally, we have information about certain types of language pathologies which give rise to different patterns of breakdown of our language skills. Data about the impact on speech of various cognitive disorders can show how a speaker's complex of linguistic abilities are organised and how they individually participate in the functioning of the whole system.

All the above material is classifiable as data only relative to a theory. So let us return to the nature of theories of language and ask what we need a theory of language for. The best way to answer this is to isolate the specific question a theory of language is supposed to provide an answer to. We need to distinguish the linguist's task from the theoretical tasks of

Chapter Eight

those working in related disciplines; even if, at the end of the day we want to relate the project of linguistics to those other disciplines. Consider the following example and a number of different questions we could ask about it. Let us suppose that Gill is having a conversation with Stephanie in English one afternoon at a quarter to two. The different interests we could take in this incident reflect the different disciplinary questions and methodologies of the philosopher, the psychologist and the linguist. Philosophers like to distinguish *how* questions from *why* questions, seeing the latter as marking the domain of their inquiries. But this contrast is too simple. I take it that the question of why Gill and Stephanie were having a conversation is a question for the depth psychologist: a question which might take on a special interest if we knew that Gill was getting married that day and was due at the Church at two o'clock. But this is not the philosopher's question. We need to probe a little further. The psycholinguist might ask what enabled Gill and Stephanie to produce and comprehend those speech sounds; where this would involve questions about the proper functioning of their short term memories, their encoding of the phonological structure of English, whether they stored their knowledge of words phonologically or orthographically, and so on. But arguably there is a question which is logically prior to this one, a question that has to be answered before the psycholinguist can answer her questions. This is: what are Gill and Stephanie actually doing? Or, rather, what is it to be speaking and understanding a language like English? This is the question that concerns the linguist and philosopher. And the priority claim is that we must characterise this activity before we can explain *how* it was brought about. Let us follow up the question of what speakers are able to do.

What both speakers are doing is engaging in linguistic activity in a language they have both mastered. Possession and exercise of this skill essentially involves knowledge. To suppose that it does not would be to find nothing odd about the remark by a character in a P.G.Wodehouse novel, who when asked at a party if she speaks Spanish, replies, "I don't know. I've never tried". This example cited by Dummett (1975b p1), can be contrasted with the same reply to the question "Can you run a mile in under five minutes?". The absurdity in the first case, but not the second, is

Chapter Eight

due to the fact that speaking and understanding a language like Spanish requires conscious knowledge, and in particular knowledge that one is speaking *that* language. In the other case, one could imagine someone chased by a tiger running a mile in under five minutes, without knowing he was doing this. But one could not find oneself speaking Spanish and not know that the words coming out of one's mouth were Spanish; not, that is, if one was able not only to speak but to understand those words. The fact that we describe both complex cognitive skills like language-mastery and complex physical skills like being able to swim as cases of *knowing how* to do something may simply reflect the fact that we were taught to do both. But the cases are very different. Even though unconscious cognitive processes play a large part in sustaining our linguistic abilities, knowledge of a language requires a good deal of conscious knowledge too, and any theory that left conscious knowledge of the language we speak out of account would be an inadequate theory of language. But for particular languages, like English, we need to account for speakers' knowledge of language in detail since possession of this knowledge distinguishes someone who is able to speak English from those who can't. It is this knowledge which determines which language a speaker speaks. So a theory of language must be a theory of speakers' knowledge of language. Let us call this the knowledge of language claim. This is the requirement that a theory of language should describe the actual knowledge speakers possess. But to assess the claim that the theory is a model of a speaker's actual knowledge we need to ask two questions:

- (1) Which aspects of speakers does the theory describe?
- (2) Which features of the theory do the describing?

Before we can start on these questions we shall need more details of the theory. A speaker's exercise of her ability to speak a language gives rise to a number of complex phenomena that can be described by highly complex principles of linguistic theory. The sounds she makes can be partitioned according to the phoneme boundaries of English dictated by phonology. Morphology describes the clustering of phonemes and the syllabic structure of words; it includes syntactic information about the

Chapter Eight

derivational history of a word from its lexical stem, such as the relations between derived nominals and their corresponding verbs (eg. criticism/criticize, destruction/destroy). Slightly larger chunks can be described as strings of words, some of which make up sentences. Different forms of structure can be described here, including different units of language larger than the word and smaller than the sentence. This is called constituent structure and without principles to describe it our theory would be unable to heed the fact that sentences composed of different word strings can exhibit common structures. At the semantic level too we must discern the recurrent items in sentence structures. For this we need semantic rules that give the meaning of lexical items and semantic principles that describe the semantic upshot of their legitimate syntactic compounding. Beyond this point, may lie the larger structures such as those described in the pragmatic theory of discourse representation theory. The possibilities of meaningful expression in a human language are traced out by the complement of sub-theories that state the principles of correct applications of these linguistic concepts together with some overarching account of their interaction.

I am assuming that the principles that describe (most of) these concepts of structure are necessary to describe the linguistic structure and content of anything as complex as a natural language like English. But it is immediately recognisable that ordinary speakers need know none of these principles (in any ordinary sense of 'know') to be able to function competently in their use of the language they describe. Ordinary speakers may consciously know that they are speaking English, or Russian, or whatever, but they do not need to know the principles by which, according to theory, they are performing such a feat. I take it that a more controversial claim would be that they do not need to know any of the concepts governed by these principles. Issues become tricky at this stage. Philosophers who accept the knowledge claim above are wont to say that speakers know the meaning of an indicative sentence because they know its truth conditions. Speakers may not have the concept of a truth condition but they have the concept of a type of situation in which the sentence concerned is true, and we represent this concept theoretically as a truth condition. But do we want to say that in all cases of linguistic

Chapter Eight

concepts belonging to the sub-theories the speakers have knowledge theoretically representable by those concepts? We can consider everything from truth conditions down to allophones. The question is how many of these notions are involved in a speaker's knowledge of language.

Because I advocate the cognitivist approach to language - one which sees the structural properties of human language as reflecting facts about the cognitive structures and operations of the human language user - I am inclined to see each of these sub-theories as addressing a different level of processing in the speech chain that takes us from physical sound to meaning. To understand the speech chain we must understand every stage of this transition in the computation of meaning from sound. This is an enormously complex task involving the imposition of many constraints on the inputs to each level: the information must be chunked and ordered under pre-specified categories, filtering out the rest as mere noise in the system. At the lowest level of input we are dealing with purely physical sounds and at the highest level we have abstract mental representations of structure. To understand the computation we must understand how sensory input of continuous acoustic signals can be analysed so as to make initial contact with our linguistic knowledge. The interface between the input and the application of phonological information is itself extremely complex. No two utterances of the sound /p/ will be the same physically. This is true even for the same individual speaker as a sound spectrograph can show. What this shows is that there is an important distinction between objective physical sound and perceived sound. Across sounds there will be perceptible physical differences: but these are not necessarily accessible to the conscious language user. The perceived difference between articulations of the bi-labial plosives, /b/ and /p/ depends on the delayed start of voicing for phonemes following /p/, absent in the case of the voiced plosive /b/. This short delay is known as the voice onset time, and it is our sensitivity to this which accounts for the difference in our perception of the two sounds. Vietnamese speakers, by contrast, have six sounds between /b/ and /p/ which English speakers cannot discriminate. This suggests different sensitivities to variation in voice onset times. Our recognition of which sound is physically realised by a speaker depends on how much tolerance we permit in what we count

as physical realisations of the *same* sound. This is important since the difference between phonemes can mean the difference between words and this can mean a difference in the perceived meaning of a speaker's utterance. And so sameness or difference between individual speech sounds depends facts about speakers' similarity judgements based on their stored information about the permitted tolerance for a sound's variable realisation. An account of what speakers know when they know the difference between speech sounds will call for a cognitive phonetics, where this will be an account of that knowledge which invokes both low-level information-processing states and high level conscious judgements. Such a theory would describe the initial stages of the speech chain from sound to meaning, and would be an indispensable part of an account of what speakers know.

This first stage of the speech chain would involve segmentation of the continuous sound signal into stable units which could be matched against our stored knowledge of phonemes and words. In fact, the whole process must relate the various minimal units of analysis at each stage: sounds, phones, phonemes, syllables, morphemes, words, phrases, constituents, sentences and possibly discourse structures to one another. I suggest that the whole task of describing this process should be tackled by constructing a computational theory analogous to Marr's computational theory of vision. We need a hierarchical model of how the human language processor makes these computations from sounds to meanings. A theory of the computation would also have to explain the relation between the levels and the constraints on the transitions between levels. Each distinct sub-theory of linguistics would provide a proper part of the explanation of the computations taking place at each levels of processing. The overall processing story would have to tell us how in comprehending a speaker's utterance, the hearer moves from a continuous sound signal, as input, to knowledge of the structured meaning of what was said, as output. Reflections on the nature of linguistic knowledge in the light of some of empirical evidence can already provide important conclusions about these connections, as I will show in a moment.

The whole computational theory of language will provide an articulated description of a part of the cognitive system at various

linguistic levels of description and explanation. Claims about the nature of information handled at those levels will depend on claims about our cognitive architecture. The cognitive architecture will engineer various solutions to information processing tasks that will constrain the information available to the system. Setting the perceptible phoneme boundaries of one's native language in the early stages of language acquisition determines what one will count as speech sounds of one's mother tongue; and this tuning-up of one's speech perceptions takes place as a result of information transfer between the level of phenomenological experience and the sub-personal processes. When the categories are stable we can say that these sub-personal settings explain our distinctive phenomenological response to a certain range of humanly produced sounds. These developments take place with respect to the early stages of speech processing but there are analogous cases we can look for in the levels above. Of course, I do not intend to give a fully articulated description of the computational theory of language processing: this is a collaborative task for linguists, psycholinguistics, computational psychologists, computer scientists, logicians, as so forth; a task for cognitive science. My aim is simply to point out the philosophical relevance of such a project and the relevance of philosophy to it; and to ask is there something radically distinct about the task of the philosopher here? Are his theories explanatorily prior to those working in neighbouring disciplines? Indeed, need he call on these empirical theories at all to complete what he has traditionally seen as a purely conceptual inquiry? This depends on what his aims are of course. But if the philosopher of language is attempting to cast light on the questions of how finite creatures can understand a potential infinity of sentences in a natural language, how they can learn such a language in a finite time, and how they can understand sentences they have never before encountered, is he not asking for an empirical theory that will explain the actual linguistic abilities of speakers? If the claim is that this not the case, and the aim is to provide a purely philosophical theory of meaning then I want to ask whether philosophers of language can afford to rest content with what Crispin Wright (1986a, p32) has called the "cosily collaborative" picture, according to which it is the philosopher's a priori task to devise theories

Chapter Eight

in line with certain key assumptions and adequacy conditions, and it is the task of the empirical researcher to determine whether or not the a priori theory reflects the actual states of speakers? (I suspect this collaboration is less cosy for the empirical researchers; for it looks as though the philosopher has nothing to learn from them except after the event.) This depends, as Wright says, on what properties of speakers the theorist of meaning takes himself to be describing. Wright thinks this is a philosophical matter: the philosopher's task being to state the conditions on a speaker's having knowledge of meaning, and the empirical psychologist's task being to tell us whether the conditions are met by anyone. I disagree. My burden will be to show that it is not always up to the philosopher alone to decide which aspects of a speaker the theorist of meaning is trying to describe. Philosophers can be and have been mistaken about what they themselves are trying to describe in their theories of language. Proper attention to the provenance of the concepts we use in constructing our 'philosophical' theories will show that they have an everyday usage and a more technical usage. Often the everyday usage simply presupposes that they can be given a more technical sense, without offering any indication of what such an account might be like. So concepts of this sort exhibit the phenomenon Putnam calls the division of linguistic labour. In these cases our ordinary use of the concepts of gold or air, and claims we might make about such things, would be given a more precise sense by the relevant range of experts. The linguistic concepts of word, sentence, grammar and language have an everyday sense and a more technical usage, and it would be wrong to assume that just because we ordinarily know which things are meaningful sentences that we know what a meaningful sentence is, or what property of a speaker amounts to his possessing knowledge of a particular sentence. The question is whether the philosopher of language can really rest content with the everyday usage of these notions, or whether he is presupposing some further account of them. The trouble with the former option is just that we are trying to give a theory of these things, and ordinary knowledge is not theoretical knowledge; it is just this admission which requires us to say what the relation is between theories of meaning and individual language users. But if we take the latter option, then I claim that prior

Chapter Eight

knowledge of empirical findings is required to determine the subject matter of the theory and to determine which states of a speaker are being described. Nobody would deny the philosopher or the AI programmer the right to invent formal systems and then to ask the psychologist whether there are any creatures whose (behavioural? neural? conscious?) states can be described by the patterns in these systems (eg. rational decision theory). But when philosophers take themselves to be proposing theories of meaning for natural languages actually spoken by some community, albeit under certain idealised regimentations of the data, they cannot merely help themselves to a rich theoretical vocabulary for specifying the a priori theory and then go on to stipulate that the particular descriptions contained in these theories correspond to, say, the patterns of behaviour in linguistic use. Heed must be paid to where we get the materials for constructing such theories: the provenance of those concepts. A proper understanding of what the significance of certain concepts is owed to may show us that by its use we are already committed to describing certain aspects of a speaker, as we saw in the case of the concept of a phoneme. The epistemology of language dictates the objects in the ontology of language, so we need to know what goes into an account of the ordinary speaker's epistemology of language. I shall now show that we need a mixture of philosophical reflection and empirical theory to say what a knowledge of the identity of individual sentences consists in. Detailed treatment of this issue will unfold throughout this chapter, but at this stage, I am merely seeking to open up a space for a computational account of a speaker's capacity to speak his native language. It would provide an articulated description of that part of the cognitive system that performs the task of linguistic comprehension (and production). This type of computational theory of language provides a partial model of the mind.

8.2) Knowledge of Syntax

Chapter Eight

I propose to begin this large scale task with an argument from syntax since this is a relatively high-level property of the sentences we utter and intimately related to the meanings that our sentences can express. The task syntax performs in a language like ours is easy to state: it puts the words in the right order. What is not so easy to state are the rules that describe what it is for words to be in the right order. But I suspect most philosophers of language simply assume that there is a story to be told of how it is done and that we can simply take this account for granted to get onto the interesting philosophical issues about meaning. This is a mistake. Only by getting the words in one of the right orders do we mean anything sensible at all, and the question of how this is done by speakers is of great philosophical importance. Philosophical interest should be aroused too by the fact that the word "grammar" is used ambiguously to mean both a speaker's knowledge or 'cognisance' of the syntactic structure of sentences and the linguistics theory of this. It is obviously a large philosophical issue what a speaker's knowledge of grammar is and what his having it consists in. We cannot simply equate knowledge of grammar with theory of grammar without saying something more. After all, unreflective language users do not know a theory of syntax for their language; or at least they do not know one in any ordinary sense of 'know'. Chomsky tells us that 'grammar' means both a cognitive structure and a linguist's theory of it. So large claims are being made by the linguist's use of theoretical terms and the philosopher reserves the right to look at those claims very carefully.

Part of this project will be to discover where facts about the grammar of a language are to be located. In saying which properties of speakers a theory of meaning describes, Dummett talks of the need for certain "linking principles" to connect theoretical notions to the linguistic practices of speakers. (Dummett 1986a, p467) As he says, these linking principles are usually implicit and unexplained by the theorist of meaning, and yet until we have an account of them it does make sense to ask whether the theory is correct or incorrect. I take myself to be giving a part of the account of these linking principles for that part of a theory of language concerned with syntax; and once again the question will be which aspects of speakers does a theory of syntax describe and which

Chapter Eight

features of theory do the describing. An account of these linking principles is of fundamental importance to the philosopher of language for only a finished theory which include these can tie together the three central strands of language mastery: meaning, knowledge and use.

Should the linking principles connect the concepts of syntax with properties of observable behaviour or should it connect them with a speaker's cognitive psychological states? Philosophers usually ignore this question by talking about the questions that arise after the provision of some suitable syntax. But they are all agreed that it has a role to play in the theory of meaning. Wright declares that theories of meaning of the kind we have been considering will have a syntactic and a semantic part. "The syntactic part", he tells us, "will divide the the atomic expressions of the language into finitely many basic syntactic kinds, and determine on the basis of that division which combinations of such expressions are grammatically well-formed." This is no mean task; and whatever properties of a speaker these theoretical descriptions relate to, it is these items which are also supposed to have meaning. Later on in the collaboration Wright imagines a two stage process of theory construction and testing. At the first stage the armchair theorist would give an axiomatic characterisation of the relevant properties of sentences; the second stage would involve the empirical researcher trying to work out how human beings could embody the information represented in the theory. But attention to the actual practice of theory construction in theoretical linguistics would show that it is in the course of constructing such a theory that we come to decisions about the aspects of speakers we are describing. The term "state" may be our term of art but it is not just up to us to specify contents in our theory of a speaker's language independently of undertaking commitments to the sort of things we are describing. By the time we have a theory of syntax of the sort Wright describes we shall have committed ourselves to describing a speaker's mental representations of structure. It will then be a further question what it is for a physical device to have mental representations, just as we can ask what it is for a physical device to run a high-level programming language and embody the structures of the corresponding virtual machine.

Chapter Eight

Not all philosophers take the ecumenical stance of Wright or show his reticence to pronounce on the nature of linguistic facts. When asked whether semantical facts are grounded in behavioural facts, psychological facts, social facts, or facts about a platonic realm, they will feel themselves able to rule out some of these options a priori and in advance of empirical research. Such thinkers believe that we can establish the claim that semantics cannot be grounded in psychology just by reflection on the nature of semantics and reflection on the nature of psychology. But I think we can afford to be rightly suspicious of these claims. Quite often, we need to know more, empirically speaking, about data and the experimental findings of a theory before we can pronounce with any authority about the nature of its subject matter

McDowell speaks of the "the match between theoretical syntax and actual utterance events" conceiving of syntax as "configurations observable in physical utterance events". But his modesty allows him to bring to bear concepts already available to us in the course of exercising the understanding the theory is trying to explain, so perhaps he has recourse to a rich notion of what we perceive when we hear physical utterances. But it appears not because as we saw above he goes on to tell us that syntactic properties are to be found among "the hard physical facts..which constrain the construction of a truth characterisation for a language actually spoken" and he conceives them as "structural properties of physical utterance-events which permit the language to be given a syntactic characterisation". (McDowell 1978, p122). So we know he thinks of syntax as describing physical facts; those facts on which we can put an interpretation. Here is an a priori decision about the nature of syntactic facts. It is also wrong. To see this we must struggle out of the armchairs and look at some empirical evidence.

To begin with, even at the earliest stages of linguistic processing before we get to syntax what we must recognise, or register, is information about the structure of phonemes in a speaker's utterance. This is information that is perceptible though not phenomenologically salient, such as perception of a voice onset time. And as we saw earlier, phonemes are not physical entities, they are variably realised as physical sound within certain tolerances. We have to have some way of registering the degrees of

Chapter Eight

tolerance in the acoustic properties of what is to count as the articulation of a /t/ or a /d/, say, especially since the phonetic realisation of each phoneme depends on the context in which it occurs. Now if we want an account of what we know when we hear that sound, can we do better than to say modestly that we hear it as a /p/ ? Yes, for we can give a full-blooded account of this information state by means of distinctive feature phonology which refers to properties of the articulation of a word which are below the level of consciousness for ordinary language users. Each informational state is a cluster of features about the sound heard, either (+/- voiced), (+/- rounding), (+/- continuant), (+/- vocalic), etc. The principled description of these linguistic phenomena draws on notions of information not conceptualised by ordinary speakers. Nevertheless, we can show that speakers are sensitive to these features. On this basis, we can argue for their psychological reality. This is not just a claim about speech perception. In fact, it would be hard to claim that our sub-personal 'knowledge' of the distinctive features of phonemes is *constituted* by the information state involved in speech perception since this information affects perception, has an effect on articulation and is productive. For example, take the phonemes that realise the -ed affix in regular past tense endings of verb stems. Consider, the three past participles, "walked", "jogged", "hated". Pronounce them aloud. You will notice that you pronounce the -ed ending differently in each case. They are realised by the phonemes, /t/, /d/, and /Id/ respectively. Consonant clusters at the end of a word never show the sequence (unvoiced-voiced) or vice versa, "walk" ends in an unvoiced consonant, and "jog" ends in a voiced consonant, so they are followed, respectively, by unvoiced and voiced phonemes; both of which realise the same morphological function. So when we perceive the consonant at the end of a regular verb stem we have to follow an instruction to produce the correct realisation. Devices which are not sensitive to the internal structure of phonemes could not make these adjustments. Also the subject has no difficulty in projecting this property to the articulation of new words, or non-words. Subjects do not have a degree of freedom to extend to new cases in any way they chose. This productivity argues for a generalisation about a level of unconscious information processing that requires analysis of input, classification, and

Chapter Eight

rule-governed output. This can only be achieved if the processor has access to structured representations of the distinctive features involved. Now it would be hard to describe these data in terms of purely physical processes since the structures we are describing are not physical structures. What we are describing is part of a speaker's cognition, and not just events that tend to happen in the normal physical course of things. These properties affect perception, and speech production, and we can be made aware of them phenomenologically once our attention has been drawn to them. What we have is an information processing story about states "that play a role in the proximate causal history of beliefs, though they are not beliefs themselves", in Stich's words "subdoxastic" states. These are not necessarily unconscious states since the state bearing information about the distinctive features of these phonemes can surface in consciousness with a little persuasion. They are however non-conceptual in the sense of not being conceptualised by the subject (until he consciously attends to them). There is still a distinction between what you now know as a theorist of phonology and your *hearing as different* phonemes which are voiced and unvoiced. However, this is a very low-level representation of information and perhaps not worth calling *mental* representation, but we can now move to analogous cases of conceptual structures higher up the processing hierarchy.

Let us turn to syntax. Although linguists of Chomsky's persuasion use the word 'grammar' with studied ambiguity to mean either an internalised cognitive structure of the speaker or a theorist's description of the properties of that structure, there is still a difference that Chomsky points to between a speaker **knowing rule R** and a theorist **knowing that rule R is a rule of the speaker's internalised grammar** (even if the theorist and the subject are one and the same person). One difference is that we cannot *express* the speaker's knowledge of R, we can only *represent* it by a theory. To adopt Wittgenstein's terms and reverse them, we cannot *show* this knowledge we can only say what it is, so we can give an account of it. So Chomsky provides a theory of what speakers know of syntax. A theory of what we know must be a theory of our grasp of that knowledge; so a theory of what speakers know of syntactic structure must

Chapter Eight

be a theory of their grasp of such structure. This is what Chomsky offers us in giving a full-blooded, rather than a modest, theory of syntax.

Consider the following pair of sentences:

- (1) Linguists are difficult to employ
- (2) Linguists are keen to learn

Do they strike one as similar? The behaviourist might well admit there was a similar pattern here. Do we perceive them as having the same structure? By McDowell's criteria we probably do. But are they syntactically similar? Consider the following related pairs of sentences:

- (3) It is difficult to employ linguists
- * (4) It is keen to learn linguists
- (5) Employing linguists is difficult
- * (6) Learning linguists is keen

Clearly, (4) and (6) are aberrant in some way, and yet we have performed the same re-arrangements in the case of (3) and (5) which seem okay. We know this straightaway and never try to produce sentence like (4) and (6) on the basis of our successes with (3) and (5). How is this to be explained? The answer is, of course, that "Linguists" is the grammatical object of "to employ" in (1) and the grammatical subject of "to learn" in (2). Moral: surface syntax is misleading. Now we must have some way of registering to ourselves the grammatical relations involved in these sentences, even if we cannot say what a grammatical subject or object is. The fanciful suggestion is that we have represented in some way the structure of (1) and (2) as follows:

- (1') Linguists are difficult [e to employ e']
- (2') Linguists are keen [e to learn]

This is known as S-level in Chomsky's Government/Binding theory (GB). This is a level prior to phonological realisation at which crucial information is recorded which has a role to play in both phonological realisation and semantical interpretation. Elements like 'e' are called

Chapter Eight

empty categories. Although these empty categories are often phonologically unrealised, some cases permit realisation. For example, the first empty category *e* in (1') can be lexically filled by the formative "one" when accompanied by the complementiser "for", giving us:

(1'') Linguists are difficult [for one to employ *e*']

Some may find it egregious to admit to such extra levels of structure and balk at the thought that we are claiming them to be somehow psychologically represented by the speaker. The mention of empty categories will seem unparsimonious, psychologically extravagant and wholly dispensable. From these examples alone I would agree. However, we can give stronger reasons than this for acknowledging empty categories in our theoretical representations of speakers' knowledge. Here are a few more.

From the postulations of two empty categories in (1') nothing so far prevents us from linking "Linguists" with the first element *e* to give the reading 'It is difficult to get linguists to employ people', and yet that is not an option for (1). That is not a meaning we would consider. So how is it ruled out? It is prevented by a structural condition of control that the understood element for 'one' never occurs as an object, and so 'Linguists' cannot be the subject of 'to employ'. Explaining why (1) is unambiguous and why certain meanings are not available to us is just as important as explaining what these sentences do mean. If not just any way of 'going on' constitutes a correct use of our language then we need to know the constraints under which we are operating. The structural conditions do not provide descriptions of our intentions to use, but constraints on how we can implement our linguistic intentions. And these constraints place conditions on what speakers of a language can be said to know. The notion of constraint has become of increasing importance in contemporary grammar, gradually replacing the notion of generativity. But before I say more about this we still need truly convincing reasons to admit the existence of empty categories.

Consider the following four sentences:

(7) Who did Bill see?

Chapter Eight

- (8) Bill sees Mary
- (9) Bill sees
- (10) Who did Bill see Mary?

I think the intuition is that (7) and (8) are acceptable, but that (9) and (10) are not. What is wrong with (9) is that "see" is a transitive verb and requires a direct object. Sentence (9) has a subject but no object to follow "see", therefore it is ill-formed. But what happens in (7) where there is no direct object but the sentence is grammatical? Clearly, the requirement is suspended. Perhaps it is optional? But then (10) shows us that there are cases where the direct object must be suspended on pain of ungrammaticality. We can see the reason for this straightaway: the relative pronoun "who" is the direct object, so when it is present, as in (10), the position following "see" must not be filled; and when it is absent in (8) and (9) it must be. Hence, the irregularity of (9).

The importance of this example is that shows why we cannot explain these data by features of simple linear order alone. It is not optional whether we have a lexical item in the final position of the verb phrase so we could not construct a context free grammar for these sentences with the following re-write rules. (Bracketing indicating optionality.)

- (i) $VP \Rightarrow V (NP)$
- (ii) $S \Rightarrow (Who) NP VP$

Neither of these rules is correct since they both permit ungrammatical constructions. The problem that confronts us when trying to write a context-free grammar is that there is a dependency between these two rules: the NP must be suppressed in (i) when 'Who' is realised in applications of (ii), and 'Who' must be suppressed when the NP is realised in applications of (i). The dependencies between these two rules must be expressed by a higher-order rule that describes not just categories and structures but relations and dependencies between them. For this we need movement rules. The thought is that "who" is a displaced element which would have started life in that post-verbal position, as we can see in the case of so-called echo questions:

Chapter Eight

(11) Bill saw who?

The usual story is that "who" has been moved by a transformation (wh-movement) that takes it to the front of the sentence.

The example also gives rise to claims about psychological reality. If we are sensitive to the conditions on structure described in a grammar adequate to represent what we know, then our knowledge of syntax is structure-dependent. The Structural Dependency Principle says that speakers of natural languages require knowledge of the relationships between words and not just knowledge of their linear ordering. Chomsky glosses it, not in terms of knowledge, but as a principle at work in computational cognition:

children unerringly use computationally complex structure-dependent rules rather than computationally simple rules that involve only the predicate "leftmost" in a linear sequence of words" (Chomsky 1986, pp7-8)

Whether one wants to call the states involved here 'knowledge', they are certainly implicated in the computational mechanisms that determine the uses, misuses, and restrictions we observe in language use. And although the computational operations themselves do not have to be explicitly represented in the system, the structures over which the computations are performed do. But Chomsky can claim more than this. Structure-dependency is a universal property of all human languages; and since a theory of language is a theory of speaker's knowledge of language, speaker's must know the structure-dependencies the grammars of their languages describe. This knowledge of structured objects does not yet entail structured knowledge; but then, if there is a mechanism for explaining our knowledge of these structured objects and if that mechanism includes structured representations that causally affect the sorts of things we do with our language, there is a strong inclination to say that this does amount to our having structured knowledge. The argument for this would be a 'How else do we come to have knowledge of something structured?' challenge. If we have knowledge of structured

Chapter Eight

objects and their structure is not a reflection of the structure of our knowledge, nor of their overt linear form, then where else does this structure reside? Platonism is one suggestion, but it is epistemologically problematic and cannot explain the generalisations about speaker's use. Why should our cognitive operations show such conformity to properties of these abstract entities? Is it not more likely that these abstract properties of structure are projections from our cognitive functionings? It is in this spirit, I believe, that Chomsky says: "language has no objective existence apart from its mental representation" (Chomsky 1972, 169). This why 'knowledge of language' is a structured state of the language user and not a speaker's epistemic relation to something independent of him. This is a very special sense of knowledge and for the moment, without saying any more, let us call it **tacit knowledge**.

Perhaps we are still not convinced that this provides a case for mental representation. But there is more to be said. The theory of GB requires that there should be an empty category in object position in (7), and that this position should be assigned the same 'thematic' role (eg. agent, patient, instrument, etc.) as the the moved element "who". This is a special kind of empty category called a trace. We can think of it as the trace left by the moved element. Its presence indicates a chain linking "who" and its trace, which ensures the sharing of certain grammatical information (such as that it is a direct object) The details are described in a part of the grammar known as trace theory. The structure of (7) is as follows:

(12) Who did [Bill see t]

Two questions arise straightaway: 'What is the motivation for trace theory?', 'Are traces psychologically real?' Let me take these in order.

To understand trace theory one must understand something about the 'architecture' of Chomsky' Government/Binding framework. I cannot hope to give an adequate exposition of this highly sophisticated theory here, so I shall simply provide a short tour pointing out the sights that are important for our purposes.

8.3) Principles and Parameters

GB is a collection of interrelated sub-theories. These sub-theories deal with the highly interactive principles of Universal Grammar (UG). UG is "a system of principles, conditions, and rules that are elements or properties of all human languages". A Universal Grammar gives the essential properties of language, and it describes an innately given part of a speaker's linguistic knowledge. The principles of UG allow for certain parametric variations within fixed limits. These are the variations between languages. Universal Grammar is not the grammar of any particular language, it describes the initial state of the language faculty. For us to acquire language we have to acquire a particular grammar determined by features of the language used in the community in which we first learn language. Acquiring language means coming to apply the innately given principles of UG to the data of a particular language and setting the values of the parameters locally according to that language.

An example of parameter setting is the Head-parameter which specifies the order of elements in the syntactic structure of different languages. In English, the lexical heads, or dominant lexical categories of the major constituents of a sentence, Verb Phrase, Noun Phrase, Prepositional Phrase, all occur on the left of their phrase structures:

(VP) carried his books to the station

(PP) to the station

(NP) man with the dark overcoat

In other language, like Japanese, the lexical heads appear to the right of the other elements in their phrasal categories eg.:

(PP) Nihon ni. (Japan in)

(VP) E wa kabe ni kakatte imasu. (picture wall on is hanging)

Chapter Eight

It turns out that languages are internally consistent and that the position of heads in all the phrases of a language can be specified as either **head-initial** or **head-final**. This illustrates one of the features of the Principles and Parameters framework: there are no construction specific rules. So instead of a long list of phrase structure rules from which we could work out the positions of lexical heads in phrases, we get a single generalisation. All the incipient language learner has to do is discover whether her language is a head-initial or head-final language and set her parameter accordingly. Children need evidence of this, but even simple three word sentences like "John ran home" will provide it. This is the head parameter.

Several other parameters exist involving the extent of the structural dependencies of movement in various languages. Languages exhibit a strict hierarchy of positions on which one can move a wh-word so as to question or relativise, like the wh-movement of "Who" in the example above. Moving from left to right, there is a strict ordering of the noun-phrase positions in a sentence which are accessible for relativising: subject, object, indirect object, prepositional object, and complement positions. Different languages stop at different positions along this list, with relativisations on positions beyond the stopping point considered as ungrammatical. Each language stops somewhere along this left to right list, but no language has gaps. This is known as the Keenan-Comrie Relative Clause Accessibility Hierarchy Parameter.

Chomsky has said that:

Ideally, we hope to find that complexes of properties differentiating otherwise similar languages are reducible to a single parameter, fixed in one or another way (Chomsky 1981, p6)

The claim would then be that starting out with UG, if we set the parameters one way we get one language, if we set them another we get a different language. This would be a strong vindication of Chomsky's claims for Universal Grammar.

8.4) Government/Binding Theory

Chapter Eight

Because of the extreme generality of the principles, more and more information is carried in the lexical entries for each word rather than by rules, as it was in transformational grammar. The lexical entries will be categorised under lexical headings, like Verb, Noun, Determiner, etc. But they will also be *sub*-categorised. So for example, verbs that are transitive will be subcategorised as having to have an NP after them. The information is recorded in theory by subcategorisation frames. These will include information about the number of argument places for a verb, and the number and position of its complements. (Note that the arguments of a two-place verb like "loves" will require it to have a subject and an object. But the complement is the argument that actually occurs in the verb-phrase itself, ie. its object)

Subcategorisation frames look like this:

v. gives [_ NP1 to NP2]

The brace tells us what syntactic contexts "gives" can occur in. They will also tell us when verbs like "give" can also occur in the syntactic contexts:

v. [_ , NP2 NP1]

Note that while "give" can occur in the second kind of context, a similar verb like "donate" cannot. This subcategorisation information is obviously interactive when several words are placed together. If they each fit into one another's slots then we have the beginnings of a sentence structure: a D-structure. There is only one transformation rule that takes D-structures to S-structures; this is "Move -a", where "a" is a variable ranging over any grammatical category. It says, in effect, move anything anywhere. What stops this overgeneralising and permitting absolute nonsense is the constraining effect of the principles. The principles interact to rule out impermissible structures. So any word-string is possible unless it violates one of the principles. In this way, GB is not a generative theory: it is a constraints theory. It says that however we arrive at word strings, if they violate the principles of structure then the result is

Chapter Eight

ungrammatical. Gaps in the data are meant to reflect the operation of a constraint. But of course, the constraints do not say we *cannot* construct ungrammatical sentences, they merely account for our finding certain (perhaps randomly produced) word strings unacceptable.

To give an example of the interaction of principles I need to mention two more notions. First, the Projection Principle; this says that information in the lexical entries should be projected to every level of syntactic representation. A syntactic structure is a projection from lexical properties, so subcategorisation information must be preserved. Next there is a part of GB known as theta-theory. This is the part that specifies the thematic roles of agent, or patient or theme for arguments of the verbs and prepositions. Theta roles are assigned at D-structure in accordance with the theta criterion:

θ -Criterion: Each argument bears one and only one θ -role, and each θ -role is assigned to one and only one argument.

Now, since syntax is a projection from lexical properties, there is a requirement that every lexical head should get the correct number of arguments specified for it in its subcategorisation frame; and every such argument should be assigned one and only one θ -role. These requirements are related in GB by a principle known as θ -marking. This principle makes the two systems interact as follows:

If X subcategorises the position occupied by Y, then X θ -marks Y

This means that subcategorisation entails θ -marking. This in turn entails a constraint on the mapping between D-structure and S-structure. Elements cannot be moved from one subcategorised position to another for then the moved element would have two θ -roles assigned to it in violation of the θ -Criterion. It would have a θ -role in the position it had before movement, and in the position it had after movement. All this entails that an argument (or complement) in a phrase structure can only move if it moves to a non theta-marked position. Only some positions are open for movement: these are known as landing sites.

Chapter Eight

Now because a verb phrase is only well formed if its head occurs in the syntactic context specified in its lexical entry, any 'moved-from' position in that context will have been subcategorised for and theta-marked. This has the upshot that the well-formed configuration:

v. see [_ , NP]

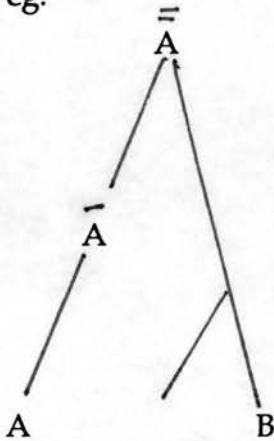
must be preserved as it is projected to all levels of syntactic representation. Once a syntactic position exists it must continue to exist at each stage of the derivation from D-structure to S-structure and to the level Chomsky calls Logical Form (LF) where information about the relative scopes of expressions is made available. After movement, a syntactic position that existed must still be available in the structure. Thus the Projection Principle entails the existence of empty categories; in this case traces.

So if, as in our example of "Who did Bill see?", the element has been moved out of the direct object position, there must be a trace left there to conform to the Projection Principle. So the sentence takes the S-structure: Who did [Bill see t]

There are other excellent empirical motivations for traces. Lexical material cannot be moved into the position occupied by a trace. The relation between a moved item and its trace is that of bound anaphora - like the relation between a reflexive pronoun (eg. "themselves") and its antecedent. This relation must satisfy certain structural conditions such as c-command. This is a relation between nodes of a phrase structure tree: a given category A c-commands another category B when every branching

structure of a maximal projection (highest phrasal category of the head's type) dominating A dominates B.

eg.



A moved item must properly bind its trace if the sentence is to be grammatical, and it can only do so when it c-commands its trace. This relation of c-command is an essential condition on anaphoric binding in general. (The Binding Principles in GB state constraints on co-reference.)

Given that traces play an integral role in the theory in concert with many interacting principles that conspire to explain the data, we must ask whether they have any psychological reality for speakers. This is part of the question of whether a grammar describes a speaker's mental representation of syntactic structure.

Traces do give rise to certain effects in speech so let us consider an argument of their psychological reality. The evidence for the psychological reality of traces concerns the phenomena of wanna-contraction. We know that people often make the following phonological contractions:

want to -> wanna

going to -> gonna

In some cases this sounds okay but in others it does not. So for example, we say: "I'm gonna kiss you" , but not "I'm gonna London". This should satisfy us that we have intuitions about such cases. Consider now the following ambiguous sentence:

Chapter Eight

(13) Who do you want to play?

This could be read in the sense of who do you want to play at squash, or in the sense of who do you want to play scrum half on Saturday, or to play the Kruetzer Sonata. It depends on whether you want someone else to do the playing or you want to play someone at a game. Now take the case of wanna-contraction:

(14) Who d'you wanna play?

Can it be read both ways? I think the intuition is that we can only understand it in the sense of who do you want to play at squash. The reading of who do you want to play the Kreutzer Sonata is ruled out. The explanation for this is shown by revealing the two S-structures for (13):

(15) Who do you want to play t

(16) Who do you want t to play

We cannot contract "want" and "to" in (16) because they are separated by a trace "t". The explanation now given is that speakers are sensitive to this empty category even though it is phonologically unrealised. This is because it is mentally represented, although phenomenologically and consciously inaccessible. This argues for a level of information processing at which speakers draw on the information recorded in a theory of syntax to guide their linguistic productions and judgements about grammaticality and ambiguity.

This is just one case where we have to mentally represent structure not disclosed in the surface of utterances; but there are many more. For instance, the syntactic phenomena of 'gapping', or verb-phrase anaphora in sentences such as:

(17) Martin had a glass of wine and so did Barry_____.

with the understood element being the constituent "have a glass of wine". To understand the sentence one needs to 'know' how to interpret the

Chapter Eight

missing element. VP ellipsis can also give rise to ambiguities of the 'pronouns of laziness' variety, as in the following case:

(18) Keith loved his cat and so did Lee_____.

Here we can either construe (18) as meaning Lee loved Keith's cat or Lee loved his own cat. But for both possibilities to be available we must have a way of acknowledging the different completions of the sentence. A plethora of cases for empty categories or 'understood' constituents emerge when we look at any grammar with the expressive power to describe structures as complex as those found in natural languages. The importance of these phonologically unrealised constituents in understanding sentences argues for the psychological reality of S-structure. The psychological reality of D-structure resides in the complex information we carry about the syntactic roles of individual words. Furthermore, transformations need no longer be seen as psychologically real operations. One can, if one wishes, merely think of them in terms of constraints on the mapping from D-structure to S-structure. The problem of linguistic knowledge is then a matter of how, given what we know or cognise about words, we can find grammatically permissible strings for which permissible mappings can be given. The principles of GB provide strong constraints on the range of possibilities.

What conclusions can we draw? We can begin by answering the opening question of this section: do the concepts of theoretical syntax describe patterns of speakers' behaviour, or do they describe speakers' psychological states? The answer must be the latter. Syntactic information is not a property of behaviour; it is systematic, structure-dependent, multi-levelled, and causally efficacious. In this way, the underlying representational structures give form to the behaviour whereby we realise the articulation of a sentence.

Of course it is open to the die-hard to claim that we can treat traces instrumentalistically, but it is a good question what causal-physical story of the observable facts will be. There would have to be a physical mechanism that explained wanna contractions and non-contractions, regardless of the presence or absence of empty categories or the purported

movements in the syntax. These purely physical explanations of linguistic production would presumably be separate from explanations of our perceptual phenomenology and how we respond to supposedly ambiguous structures. I fail to see what any such purely causal story would look like since there is no generalisation to be had at that level. The only answer an instrumentalist about the top-most level of description could give of what make the data take the shape there described for them will be in terms of 'whatever it takes' to get those data to conform to the instrumental generalisations. This is at best a little thin, but at worst it ignores an account already available to us at the irreducible cognitive psychological level of description, intermediate between the physical and the folk-psychological levels. At this level we can frame testable predictions about many other items of data. The computationally minded linguist is producing generalisations, the physical scientist is not. The linguist can hold onto the attractive idea that this mentally represented information about syntax is neutral as between comprehension and production and is called upon in both. If he does not fall into philosophical error in making such claims, then *prima facie*, he has a testable, well-supported hypothesis about how we do what we do with our language.

There is a possible confusion at this stage that must be avoided. These psychological explanations are not purely syntactic. The content of the relevant states of our linguistic processing system are about syntax, and the representational structures represent the underlying grammatical information with which we gloss behavioural acts of utterance. To account for them we need a content-using theory of these cognitive states and a psychological explanation of how they affect linguistic output. An important part of these outputs is that derivations of syntactic structures, however they are computed, emerge in fully conscious grammaticality judgements, including judgements about shared constituents, and, I would add, they give rise to such non-conceptual states as our hearing of sentences as structured, which depend crucially on particular phenomenological experiences. The content-using psychological explanations of the causes of these experiences refer to internalised structures of the cognitive system that do not require those structures to

Chapter Eight

be phenomenologically or conceptually available to the person whose states they are.

8.5) Syntax and the Three Stances

What do each of our protagonists make of this. Let me begin with Dummett and his Descriptivist position. The descriptivist believes that there are facts of the matter about our knowledge of language: something that knowledge consists in that exists independently of its theoretical characterisation. So what is Dummett to make of the claims for syntactic knowledge?

He should approve of Chomsky's full-blooded account of that knowledge; but he will frown upon the postulation of mental representations and unconscious psychological states, fearing some threat of psychologism. But the first thing to point out to him is that this is not psychologism in the sense summed up by McDowell as:

...a conception according to which the significance of others' utterances is a subject for guesswork or speculation as to how things are in a private sphere concealed behind their behaviour (McDowell 1981, p225)

The psychological states which constitute a speaker's grasp of syntax are neither private nor accessible only by introspection. They are empirically discovered and described facts which can be tested for by consulting all the types of evidence mentioned at the beginning of this chapter. Nor are they private, since others can know more about a speaker's internalised grammar than he does. Nor are they infallible; speakers can be mistaken in speech or judgements about their own language and brought to recognise this. Mistakes can be due to the fact that the language faculty is just one component of the mind that interacts with many others that are involved in the production of behaviour. Computational models provide a thoroughly perspicuous account of what it is for a physical system to perform some of the operations described,

they provide a respectable scientific theory of those aspects of cognition. What these cognitive theories do share with psychologism is the claim that products depend on processes, not just for their coming into being, but for the characterisation of what they produce. The syntactic facts about an utterance depend on how it was produced and arrived at. If an utterance is produced by a speaker who lacks the psychological organisation of a human language user, and cannot represent the structure to himself in the correct way, then he will not have produced or recognised a sentence of our language. On behavioural grounds alone we can be mistaken about who is a language user and who isn't, as the Turing Test for computer simulations show. To suppose that there could be a device or a creature who had some other way of producing speech behaviour which, for a sufficiently large range of cases, was not differentiable from humans' use on behavioural grounds is not to say that it is indistinguishable or identical to us in respect of knowledge of language. We are simply wrong about what we read-into its behaviour, imbuing it with a syntactic gloss it does not deserve. When two systems, or speakers are behaviourally indiscriminable this does not entail that they are cognitively indiscriminable. With a computational parser we can strip it down and see whether it is actually drawing upon the structures of our grammar. The same is not true in the case of a human being, or human like creature, but we have the sources of evidence mentioned above. If the claim is that there could be one of us that operates as we do but doesn't conform to our internal structuring (the cognitive structure described by the theory) then we are owed an account of how this remote logical possibility could be properly imagined. How else could the speaker do this? Could order really arise out of chaos? Surely there has to be some order at the sub-bedrock level that explains the simulator's ability to go on producing and responding to speech as we do; and we already have an account of what this organisation looks like. To consider this possibility as a serious challenge we are owed some detailed account of it. Inner processes stand in need of outward criteria.

Objecting to Chomsky, Dummett says:

Chapter Eight

The difficulty is that we have no idea what structure and character knowledge, conceived as an internal state, may have, apart from the structure of what is known. (Dummett 1981b, p6)

But that is precisely what Chomsky is giving when he speaks of providing a characterisation of the speaker's internalised grammar at some level of abstraction from physical mechanism. This is what he calls the speaker's I-language: "some element of the mind of the person who knows the language, acquired by the learner and used by the speaker hearer" (Chomsky 1986, p22) The term "knowledge of language" refers to a speaker's state of having an I-language. According to Chomsky, there is no good use for the notion of an E-language, in the sense of "external" or "extensional" language. What we are interested in are speakers' I-languages. Although this is an idiolectic notion, it does not involve overtones of a private language. It is investigable by others and perhaps its internal structures may be better known to the linguist than to the user. Nor need there be any fear that the contents of our thoughts and meaning will be seen to be constituted by materials internal to these cognitive mechanisms. The issue of internalism and externalism is largely orthogonal to the ones I am discussing here. We are simply giving an account of part of what is involved in a speaker's giving some linguistic significance to a heard expression in a particular situation; and for this we need to look at "components of the mind/brain that carry out the process of interpretation, abstracting from the complexities of the world of experience" (Chomsky 1989, p9). This is not to suggest that other features of his world, or his experience, do not enter into his understanding of an expression. We are simply studying the grammatical part of this process. It is also important to note here that these are accounts of facets of individual idiolects that generalise across individuals and so in that sense belong to the field of Individual Psychology (as explained above). These common features of speakers belong to the theory of the genetically pre-coded Universal Grammar:

Universal grammar.....specifies the class of systems that a person would acquire in an ideal homogeneous speech community. We may think of it as a function that maps experience in such a community into a state of the language faculty. (Chomsky 1989, p9)

Chapter Eight

The formal properties of a speaker's I-language, which include the structural descriptions assigned to sentences and expressions are characterised by a linguistic theory. Chomsky has shown (as we saw above) how "general and invariant principles of the language faculty interact, sometimes in intricate ways, to determine the form and meaning of expressions." (Ibid., p12) Chomsky's formal theory of grammar is an elaborate account of the structure of what is known, postulating different fundamental levels of representation at D-structure, S-structure and LF. But he is also telling us what the knowledge formally described in terms of these rich and elaborate arrays of structure must consist in:

Each of these levels stand at an interface of the computational system and some other system of the mind/brain. (Ibid., p13)

each structural description that a grammar (in either sense) assigns to an sentence and its expressions

...consists of a representation at each of these three interface levels. (loc. cit.)

and

Each element of these representations must be licensed by an appropriate relation to the external system at the interface.(loc cit)

from which Chomsky concludes that:

...the formal and semantic properties of linguistic expressions and a network of relations among them are substantially determined by the human language faculty itself. (loc. cit.)

If we add "substantially but not entirely" this should assuage Dummett's worries. Of course there will still be a need for further empirical research to fill out the details of how the brain embodies a linguistic processor with these properties. But that is the task for a computational

Chapter Eight

psychologist and neurophysiologist to work out between Marr levels 2 and 3 on Marr's hierarchy of levels. Levels 2 and 3 are the level of the algorithm and representations involved in the computation and the hardware realisation of these, respectively. On the hierarchy, these details are worked out once we have some level 1 theory of the logic of the computation being performed. So a level 1 theory of language could be an account of what the competent speaker can do, what judgements he can make, and so forth.

So far I see no reason why Dummett should object to this account on anti-psychologistic grounds. The position I want to make room for here and which I see as quite compatible with much else in Dummett is nicely encapsulated by Tyler Burge when he says:

Languages depend on the experiences, usage, and psychological structures of individuals. (Burge 1989, p176)

In asking what a speaker knows when he knows how to speak a language and what his having that knowledge consists in, I see no reason to think that Dummett can provide a satisfactory answer by appeal to anything less wide-ranging.

What other objections does Dummett have? We may recall his worry that a good deal of conscious knowledge is required for knowledge of language. Indeed it is, as Chomsky would agree, since it is our conscious knowledge "that expression such-and-such means so-and-so, or consists of certain words and phrases but not others" (Chomsky *ibid.*, p5) that he is trying to explain. The question is how much of our syntax is 'visible' at the conscious level?

Dummett has pointed out that from the fact that ordinary speakers typically cannot formalise what they know about language, it is wrong to conclude that their knowledge must be unconscious. The best we could say is that it is non-verbal. So perhaps Dummett would advocate conscious but non-verbalisable knowledge of syntax. This "unverbalisable" condition seems very strong because surely the theorist can verbalise his knowledge. But the reply might be that he could not verbalise his unreflective knowledge because this consists in his having certain phenomenological experiences of an auditory or visual kind; and

he could only represent this obliquely and theoretically, but could not express it. This suggests that the character of these experiences is non-conceptual, and in fact this would be confirmed by Dummett's language-first approach to our possession of concepts. For if it was conceptual the speaker must be able to express those concepts linguistically. So if it is unverbalisable information then it is non-conceptual. What is it like have such an experience of syntax? I think the best we could do is to say that it would occur when we had experiences of hearing a sentence uttered *as* structured as opposed to knowing *that* it was structured. We can think of this non-conceptual experience as analogous to "hearing a sound as from the left". These experiences of the ordinary person would not require him to conceptualise the structure (or the direction of the sound). But what we need to know is whether the experience of hearing a sentence as structured amounts to grasp of structure? But can we go that far without this entailing the subject's being able to conceptualise the structure heard? Perhaps all that can be claimed here is that he has a structured experience of grasping something linguistic. But is an acoustically structured *grasping* of a sentence a grasp of its structure? It is just not clear how structure in the perception of speech makes available to the subject of the perception a detachable syntactic property of the sentence itself; a structure to be found in other bits of speech and which guides linguistic production.

I cannot explore this any further, but the suspicion is that even if a clear notion of 'hearing a sentence *as* structured' can be made out, it does not seem to give us enough sight of the syntax. For instance, traces are not phenomenologically accessible (heard), nor is the relation of c-command, nor is the tighter syntactic configuration of government, and yet these notions all play a key part in securing generalisations in Chomsky's highly explanatory paradigm. What is Dummett to say about them? How much syntax can he reserve for unverbalisable conscious thought? Not enough, one suspects, to explain all our conscious and conceptualised intuitions about the data; nor enough to substantiate the incontestable psychological reality constituent structure has for speakers. This last claim can be confirmed by eliciting speaker's intuitions about the positions where they can interpolate parenthetical items like "I think" in whole sentences. It turns out that they can only insert it at major constituent

boundaries. This conscious reflection experiment reinforces the argument for unconscious recognition of constituent segments in acoustic perception where subjects are asked to say at what point in a sentence they are listening to they hear clicks occurring on the tape. Mostly all subjects tend to re-locate the clicks they hear to the constituent boundaries in a sentence. (Fodor and Bever 1965).

Since it would be hard for Dummett to say how much syntax is available phenomenologically, we might ask another question; namely, how much syntactic structure does Dummett think we need in theorising about language?

We saw above how knowledge of basic grammatical categories, like when a word is a noun, matters to Dummett, along with information about their syntactic compounding. But this information depends on higher-order principles concerning the relations between those categories. Already a substantial body of knowledge is involved and it is far from obvious that Dummett can preserve it all at the (personal) level of experiential states with unverballed contents. Surely he must concede that some of it belongs at the sub-personal level? However, he is reluctant to do so, and this is certainly not what he means by "implicit knowledge". He would not consider implicit knowledge of syntax to be a matter of sub-personal psychological states so much as a theoretical description of certain complex abilities of speakers manifested in their observable use of the language. If their abilities fit the requirements and shape of the theory then we may describe those abilities as constituting implicit knowledge of the information the theory describes. But not every externally imposed description of these abilities represents them correctly. So what restrictions can we put on them? One restriction might be that our descriptions should not stray too far from observable behaviour, but in the case of syntax we see two good reasons why they should. The first is the innateness claim and involves the poverty of the stimulus argument. There is just not enough in the observable behaviour of speakers to uniquely determine a choice of one correct grammar for their language. But if we allow the evidence to range beyond their behaviour, to comparisons across languages, to psycholinguistic data, to intuitions we can construct theories that are descriptively and explanatorily adequate.

Chapter Eight

But these descriptions are not recoverable from observation of the data available to the child, and are vastly more complex than anything the child could discover by making successive inductive steps. The best explanation for the child's acquisition of language is that he already possesses certain information which he automatically brings to bear on the data. The data he is exposed to are said to trigger the growth of his linguistic competence. If we specify some initial principles we can tell an empirically well-motivated story about the acquisition of a language, which requires the learning of a vocabulary and the setting of parameters *in situ*. It is this innate structuring in which a person's 'knowledge' of the principles of universal grammar consists. The second reason for diverging from behaviour is that we cannot capture generalisations without resorting to underlying levels of syntactic organisation which could not be read-off that behaviour.

What Dummett needs to embrace is the idea that both personal and sub-personal level states feature in explanations of the properties of linguistic expressions and what the mature language users' knowledge of syntax consists in. These appeals to sub-personal states in partial explanations of our linguistic abilities are quite compatible with Dummett's outlook, for mental representations of things like traces although unobservable entities are 'possible causes of change' in a speaker's behaviour and so therefore legitimate posits. (cf. Dummett 1973, p490-1)

I suspect that the only thing that prevents Dummett from taking this step in the case of syntax is his prevailing view of language as a skill which conflicts with Chomsky's view of language as an internal structure. If we call the complex skill of using and understanding a language, a speaker's **language mastery**, and we call his knowledge of language his **linguistic competence**, then the dispute between Dummett and Chomsky is a matter of whether linguistic competence consists in language mastery, or whether linguistic competence consists, in part, in the speakers possession of sub-personal and personal level states that are causally responsible for his language mastery. If Dummett thinks that a theory of linguistic competence is just a "systematisation of facts open to view" (Dummett 1981, p6) then the case of syntax proves him wrong.

Chapter Eight

His other worry will be about manifestation. Dummett's allegiance to Wittgenstein commits him to the Publicity Principle which we can formulate like this:

All that can be communicated is all that we can be known to mean;
and all that we can be known to mean is all that we can mean.

This principle does not have to be cast without reference to our means of knowing and the cognitive embodiment of syntactic information. But Dummett thinks that what knowledge of language consists in either is, or coincides with, what manifests that knowledge in observable use.

If these concepts of syntax are not manifested in our use then what makes us think they apply to cognitive structures we have in our heads? There are two possible responses here. Either they are manifested to one who has the relevant cognitive equipment (that of human language user): that is there is no recognising that someone is saying that sentence without sub-personal reception of the correct structural description for that surface string conceived as the utterance of a sentence of our language; or the other story we could tell would simply broaden the notion of manifestation to possession conditions. This is quite compatible with Dummett's requirements that we should be prepared to say what a speaker's having knowledge of his language consists in; that it should be a full-blooded account; and that whatever it does consist in the fact of someone's having that knowledge is epistemically confirmable. This can be achieved with the one important alteration that our being able to recognise the facts about someone's having knowledge of his language just by observing his use of it would not be due to what was observable on the surface of his behaviour (and thus being duped by Turing-Test worthy AI programs that simulate natural language conversations), it would be due, in part, to our having the internal configurations - facts discovered and described by the theorist - that enable us to put a construal (possibly mistakenly) on another creature's speech. Of course, if we began to doubt that the other was making sense in the way we do there would be ways to investigate this further, as we can see in the case of language pathologies. Facts about linguistic competence lie deep as well as on the surface, and it is in virtue of satisfying these facts that someone is able to do what

Chapter Eight

constitutes his having mastery of the language. The proper construction we are able to put on an utterance we hear in our language depends not just on what falls within our cognitive environments but also on the character of those cognitive environments: it is this which differentiates us from the non-language user. The notion of a cognitive environment will include information to which the speaker's cognitive system is sensitive as well as what is phenomenologically salient. The first will be partly responsible for the presence and character of the second.

Dummett's reluctance to acknowledge any underlying levels as entering into explanations of language mastery has a great deal to do with his conviction that meaning is use. If one see things in that way where else could meaning (and grasp of meaning) reside than in (grasp of) a pattern of observable use? Dummett drives not only meaning, but also knowledge of meaning into the public domain of observable use. And therefore a theory of meaning which serves as a theory of our knowledge of meaning is entitled, he believes, to focus on aspects of our use of the language and ignore substantive details about our cognitive states:

A theory of meaning of this kind is not intended as a psychological hypothesis. Its function is to present an analysis of the complex skill which constitutes mastery of a language, to display, in terms of what he may be said to know, just what it is that someone who possesses that mastery is able to do; it is not concerned to describe any inner psychological mechanisms which may account for his having those abilities. (Dummett 1976, p70)

There is something right in the thought that meaning should be revealed in behaviour, for where else could it be revealed if it is to be recoverable from observing people's ways of behaving? This is all that they can show us. But what matters here is what gloss the linguistically competent speaker can put on that behaviour. I am reluctant to conclude with Dummett that the knowing minds who do the observing, and who transmit and receive linguistic messages, have themselves to be flattened out onto that observable use. How does the behaving subject know of the significance of its own and other's behaviour? By interpretation? This does not seem a very rewarding answer. Instead I suggest that we look closely and in detail at the mind that inquires after other people's

Chapter Eight

meanings. We can only recover from behaviour what others reveal in it if we are sufficiently equipped as language users to recover it. For that reason a substantive account of our linguistic competence should look at the antecedent facts about our ability to retrieve syntactic and semantic information from, and embody it, in linguistic behaviour. Meaning can be revealed in use, but those uses are behavioural manifestations of cognitive capacities and can only be fully appreciated as such by those with sufficiently similar capacities. Meaning is still epistemically available in observable behaviour. But it is the cognitive characteristics of the observer that enable him to know it.

However, Dummett supposes that the facts about our linguistic competence, not merely our language mastery, are available on the surface of our behaviour. It is this view of meaning and knowledge as fully present in use that gives rise to the following suprisingly Wittgensteinian claim:

If a Martian could learn to speak a human language, or a robot be devised to behave in just the ways that are essential to a language-speaker, an implicit knowledge of the correct theory of meaning for the language could be attributed to the Martian or the robot with as much right as to a human speaker, even though their internal mechanisms were entirely different. (Ibid.)

Dummett needs to make clear what is meant here by internal mechanisms, for at one level, namely the physical, this is true. However, it is highly improbable that a sufficiently adequate syntactic description of the use of a language like English could be given without explicit commitments to the internal structured states of its speakers. The claim Dummett is making is that if any two speakers are sufficiently similar in their behavioural properties then they count as speakers of the same language. We have seen reason to think this is false in the case of the syntactic properties of their language. So what we need is to define an equivalence relation on "internal mechanisms" in terms of which we can define preconditions for a speaker speaking a language which satisfies that syntactic description. In this way we should be able to say which differences in internal mechanisms were irrelevant to one's linguistic abilities and which ones were essential. The relation we want here is

between a grammar and a speaker, and the one I have in mind would be this: for a grammar G and a speaker S , if S speaks a language described by G then there is a causal systematicity in the cognitive processes of comprehension and production in S with respect to the patterns of inputs and outputs described by G . Two speakers could qualify as speaking the same language only if their cognitive processes are causally systematic with respect to the patterns laid down in G for the structures of sentences. This is an internal constraint on the cognitive architectures of speakers and not merely a behavioural requirement. It does not commit one to the claim that the internal mechanisms have to involve states that explicitly represent the principles of grammar, but it does insist that if those principles are to be followed by computational mechanisms that produce the correct linguistic inputs and outputs they will have to involve syntactically structured inputs and outputs for those processes to engage with.

Of course things will be a little trickier than that since no two speakers will share exactly the same grammars; the grammars of idiolects shows marked variations. But the principle is clear enough. To determine the I-languages of individual speakers is a matter of hypothesising the pattern with respect to which their language faculty is causally systematic. This will probably be causal systematicity with respect to the patterns that relate utterances and the individual speaker's judgements. The syntactic structuring these generalisations would require in the inputs and outputs would have to refer to sentences' S-structures, at which information about empty categories and projected lexical properties are represented. There should be enough overlap of idiolects amongst a community of speakers to abstract general properties of the grammars that describe their individual I-languages to find out what speakers of that public language need to know, and what states they have to be in to be described as speakers of *that* language. Notice, incidentally that Chomsky thinks that very little empirical weight attaches to the notion of a shared language (an E-language) but he comes to this conclusion for very different reasons from those that lead Davidson to abandon talk of languages.

Now we have an answer to Quine-Lewis type worries about how we tell whether a community (or its members) speak a language for which G

Chapter Eight

or G' is the grammar. The answer can be given in term of the equivalence relation between grammars and speaker's states. But this is not a matter of behavioural differences, it is a matter of internal cognitive structuring.

I see no reason why Dummett cannot take this point and accept a corresponding revision to the notion of linguistic competence. Of course, the result will mean that the "cosily collaborative" picture that Wright suggests will be less cosy than Wright thinks. An account of what knowledge of language a speaker has will involve empirical research at the theory construction stage. But it is none the worse for that. It would be the right move to make here in the face of what a full-blooded account of our grasp of syntactic structure commits us to.

What of syntax and interpretation theory; can the Davidsonian expect to be so easily comforted? I think that the answer must be no and that the interpretation theorist has every reason to be discomfited by the fact that his best total theory of a speaker's behaviour cannot be reconciled with what we know empirically about our grasp of syntactic structure. Let us look at how this incompatibility emerges?

On Davidson's picture, what every speaker knows is what every interpreter knows, and what every interpreter knows can be stated explicitly in interpretation theory. So any part of what the speaker grasps must be accounted for at the most explicit theoretical level. But what resources can the interpretation theorist call upon? We have a pretty good idea by now: the raw behavioural data as physically described, the a priori normative principles of interpretation, and the two specific constraints on truth theory: systematicity and the propositional attitude constraint. A theory of language constructed on this basis can be seen as:

....warranting systematic imposition of interpreting descriptions on the range of potential behaviour, which constitute speech in the language, thought of as describable, in advance of receiving the interpreting descriptions, only as emission of noise. (McDowell 1977, p147)

But the concepts available to the interpreter, in constructing his interpreting descriptions, must be the "everyday linguistic and semantic concepts" that are "part of an intuitive theory for organising more

Chapter Eight

primitive data", according to Davidson. (T&I, p143). We have seen enough now, I think, to know that this course is hopeless: the concepts of S-structure, empty categories, c-command and so on are not everyday and intuitive concepts. What is more, there is every reason to think that to describe patterns of behaviour as complex as a natural language demands that we resort to something with as least as much expressive power as Chomsky's Government/Binding theory. Even if one does not subscribe to this theory, the only alternative paradigms that approximate to meeting the descriptive adequacy conditions for natural languages, namely, Lexical Functional Grammar and Generalised Phrase Structure grammar, both employ the notion of traces and other empty categories. Structure-dependency is an absolute property of language and to capture this we need to resort to means of description that go beyond "the relevantly similar sound-patterns of the speaker". Phonetic properties and surface structure are misleading as to interpretation, and the only way round this is to apply to behaviour concepts that go far beyond the "everyday" and "intuitive".

The trouble is that Davidson considers two realms about which we can make constitutive claims: the mental and the physical. When we offer descriptions about one or another of these domains he speaks of us having to stand ready to adjust our concepts to the constitutive principles of the physical in the one case and the constitutive principles of the mental in the other. But what he cannot tell us is where we should place the concepts that feature in the constitutive claims of linguistic theory. Claims about knowledge of syntax are neither matters of reasons nor causes. He only speaks in a vague way about a speaker's linguistic abilities; but he is, at least, forthright about the options on offer here:

Described psychologically, a speaker's linguistic ability is a complex disposition. Described physically, it is not a disposition, but an actual state, a mechanism. (T&I, p255-6)

Davidson is deliberately blind to any level of description between these. But we might say he simply gets the first one wrong. Being blind to the level(s) of cognitive organisation, he misconceives the internal mechanism as physical and *on that basis* concludes that empirical

Chapter Eight

investigation could turn up no "law-like correlation between the workings of the mechanism and speech". Perhaps this is true, but when we turn to cognitive mechanisms, which can be described and examined at a level that prescind from their variable realisations in speaker's physical states, we find increasing evidence of some such law-like correlation. Unable to recognise this level, Davidson rightly shuns the search for "the physical correlates of meaning" but now must hope to find everything he needs to describe language on the surface of behaviour seen through the eyes of the interpretation theorist. But as we have already discovered, the theory of syntax is not a theory of behaviour. It is a theory of the internal workings of the mind, at some level of cognitive organisation remote from the ordinarily intentional level, and at some remove from the details of physical implementation.

If the concepts of formal syntax characterise part of our knowledge of language and Davidson cannot admit these notions as embodied in the substantial cognitive states of speakers, then he is forced to leave these notions out of his claims for language and our knowledge of it. The result of this is disastrous. He claims that radical interpretation begins at home, and that we can find out whether someone is speaking our language only by interpreting them; where "interpretation" is used here in the sense of the encompassing project. Davidson thinks this shows why "an empirical theory of meaning does not stand alone" and why the linguistic assignments we make must be squared with a theory of the speaker's beliefs, and why change in the meanings we ascribe to him, must lead to revisions in our attributions of belief. Now given the total evidence available Davidson is forced to admit to this:

It is not strange that we can take the same person to be speaking different languages, provided we can make compensatory adjustments in the other attitudes we attribute to him. (T&I, p239)

Unlike the attributionist's position, the cognitive theorist of language has got substantive and detailed proposals about the language a person actually speaks. The case for syntax will show Davidson that there are less degrees of freedom than he thinks. Noise emissions are correctly re-described as linguistic actions not just by reference to a speaker's

Chapter Eight

propositional attitudes but also by reference to the syntax of his language. But with just an a priori theory of beliefs, desires and meaning, given empirical application by its imposition on physical behaviour, Davidson cannot accommodate this very important fact: Davidson's minimalist project has simply nothing more to offer.

The only alternatives to minimalism for the attributionist are instrumentalism and modesty. As I have said, instrumentalism can tell us nothing of how these important generalisations are sustained in so many cases, and cannot extend to new cases when we describe the phenomena at that lower level. If the causal account of our speech and understanding must draw upon the syntactic structures I have mentioned, and must explain not only what is produced but why certain options in linguistic behaviour are ruled-out or absent, then it cannot avoid resorting to the rich descriptions of cognitive structures and mechanisms that the Chomskian alludes to. This is an empirical matter.

McDowell, who has provided a stout defense of modesty in many places, would have to admit that the concepts appealed to in the structural descriptions given in GB cannot be presupposed as everyday concepts antecedently available to the ordinary language user; nor do they have to be antecedently grasped by anyone who can understand the theory. GB is a full-blooded theory of syntax which provides one with explanation of the concepts involved! But it would be hasty on this basis to conclude that McDowell falls foul of the same meagre resources objection as Davidson. For there are at least two more ideas that McDowell brings to his discussion of language understanding. One has to do with the claim that the ability to comprehend heard speech is an information processing capacity, the other has to do with the distinctive phenomenology of language understanding conceived as a kind of direct perception. Let us look at both of these.

Despite declaring a preference for the resources of ordinary intentional psychology nearly all of the philosophers under discussion here admit the existence of an actual mechanism in the speaker that carries out the task of processing and producing speech; perhaps Wittgenstein is the only exception. In describing such a mechanism in terms of an information processing capacity it gives rise to, McDowell shows more sensitivity to

how things stand with the speaker than is shown by Davidson in his reference to purely physical mechanisms. We saw that with only rational psychology or physics to turn to, Davidson has difficulty accommodating syntax. And yet, for all his talk of understanding as an information processing task, McDowell faces similar problems. The facts about syntax will have to be isolated either by the causal explanatory story or found in the rational explanations of linguistic behaviour. Whatever takes place in our linguistic information processors they can be described in only one of these two very distinct ways. The input to output transitions are re-described at the higher level in terms of an interpretation theory. So either the syntactic concepts are intentional concepts within the grasp of the interpreter already, or else it has to be shown that they are just elementary rearrangements of the physical facts, phenomena that arise as a matter of the way the world works. The trouble is that if syntax is relegated to this lowest level of causal interactions then it becomes hard to explain the information contained in the mental representations of S-structures, not every part of which is a physical property of perceptible sound-patterns. Also, it becomes hard to explain how these facts about physical structuring give rise to the phenomenological facts about syntactic structure - hearing a sentence as structured, or showing the relevant sensitivity to the insertion of lexical material, or (mis-)perceiving the locations of clicks - which surface in a speaker's experience. The thought would have to be, that somehow, or other, these early processing mechanisms operate to put us in a state of awareness in which our rational activities can begin to get to work on aspects of the observable behaviour of others, presenting it as being organised in a distinctively linguistic way.

But this will not do. In the official story, the two types of explanation are kept distinct: there is no psycho-physics of speech processing. Either we must consider the information processing story in the hearer without benefit or illumination from the intentional notions; or else we must stay within the domain of interpretation theory that re-describes these lower level facts by imposing higher level concepts on them. And here is the rub. Interpretation theory by itself is supposed to effect the transition between the lower and the higher levels; it alone can justify our

Chapter Eight

imposition of semantic and psychological concepts on behaviour, physicalistically conceived. In interpretation theory any evidence for making intentional descriptions must itself be described in such a way as to render it part of the constitutive claims about meaning or mental state ascription. Facts about information processing seem quite irrelevant.

So what can the appeal to information processing show, and what relation is there between the processing capacity and the high level intentional theory? As we saw above, McDowell envisages something like the following. Subjects who are party to a linguistic act of utterance receive certain low-level information, impinging on their perceptual apparatus, such that the information is shared across their perceptual fields, or overlapping cognitive environments (ie. what is cognitively as well as phenomenologically salient): certain further information arises for those of them who understand the spoken language. We need a theory of language which will explain how that noise is comprehended by native speakers.

McDowell says of this:

Such a theory, then, would have the following deductive power: given a suitable formulation of the information made available to both the possessor and the non-possessor of the state of understanding on any of the relevant potential occasions, it would permit derivation of the information which the possessor of the state would be distinguished by having. The ability to comprehend heard speech is an information-processing capacity, and the theory would describe it by articulating in detail the relation, which defines the capacity, between input information and output information. (McDowell 1977, p147)

This is on the right lines and presciently anticipates Marr's hierarchy of levels. On McDowell's reading of the situation level 1 theories would be explanatory prior to, and one-way independent of, theories at level 2. So this provides another example of the "cosily collaborative" picture of Wright.

But given that this is a level 1 theory of what is actually achieved by speakers' linguistic processing, we need good reasons at level 2 to believe that they are indeed computing the function identified at level 1. Level 2

Chapter Eight

psychological theories describe a range of algorithms for computing the function specified at level 1. What is needed from McDowell is an independently convincing reason to think that speakers are processing the structured contents he would be inclined to ascribe to them. But what would show this? When McDowell speaks of "articulating in detail the relation, which defines the capacity" he is not talking about a psychological explanation in terms of the computational mechanisms in speakers. He is talking about how we should specify the function-in-extension that defines a set of ordered input-output pairs. The "articulation" McDowell has in mind involves specifying the logic of this relation in terms of a structure-imposing deductive theory. Why should we expect these details of structure to belong at level 1, how are these claims for the structure of an actual speaker's knowledge to be substantiated, or better, justified *at that level*. In addition, we also need to know whether syntax belongs at level 1 too, or if McDowell intends to relegate it to the details of the sub-personal processing story. If it is the former, we need some level 1 justification for the concepts found there, and some way to integrate them with the normatively governed notions of belief, desire and intention. If it is the latter, McDowell seems to believe that this can be no more than a story about brute causal mechanisms. But if that is so, the suspicion will be that the interpretation theory whose task it is to conduct the transitions between the levels of mere noise and meaningful speech will be deprived of the concepts of syntax it needs to complete the theoretically insightful re-descriptions of a speaker's behaviour. Neither of these possibilities look very promising.

I suspect that it becomes very important to know which bits of sound emission the information processing capacity goes to work on; ie. which of the ambient sounds do we process as speech sounds, by successfully returning a linguistically useful identification as output for an acoustic signal as input. For if mastery is information processing, it is processing dedicated to a proprietary domain of inputs. But since McDowell has made it the task of interpretation theory to negotiate every step of the transition between levels; that theory alone must identify the sounds it includes in the range of behaviours which it tries to make sense of. But not every bit of behaviour produced verbally can count as speech. So the

Chapter Eight

interpretation theorist need not strive to incorporate all of a speaker's mouthings in a rational and consistent pattern of his meanings and attitudes. Certainly the interpreter needs to know which speech behaviours are intentional acts. But before this he needs to know which behaviours count as speech. But without empirical investigation of the inputs and outputs it seems highly problematic to suppose that interpretation theory should state in advance what it is likely to be able to treat as data for a linguistic theory. But perhaps this point is only telling on the reading of the project of radical interpretation which regards it as having a heuristic value, in showing us how the facts of meaning can be grounded in (supervene upon) the facts about behaviour. On this reading, it shows us no such thing unless it tells us how it selects its inputs for conceptual incorporation.

This is an important issue for McDowell, for if there is such a thing as a speaker's being in accord with, or going against, a rule on any occasion in his use of his own language, then we cannot simply try to make any way of going on consistent with our erstwhile interpretation of him. But what are the constraints here? If there is something about a speaker using a language which makes matters of use correct or incorrect, then one should not attempt to revise one's interpretation in the case of all variations and departures from it. I would want to claim that our mastery of syntax provides one type of condition on what it is to accord with or depart from the use of one's own language. To depart radically from this aspect of one's own language is to cease to be causally systematic with respect to the syntactic patterns described in the input-output pairings of utterances and grammatical judgements. This amounts to loss of grasp of their structure and grammatical relations. Aphasia patients may be diagnosed in just this way. But it is not obvious how an interpreter is to explain this strand of underpinning in linguistic competence.

If the concepts of syntax are among those which guide the interpreter in finding interpretations of speakers, then we have to be told how the interpreter's understands those concepts. Modesty is not enough for unlike the case where meanings can be assumed to be already in the grasp of the language user being described, there are no rational explanatory grounds for attributing the concepts of theoretical syntax to ordinary

Chapter Eight

thinkers. Unlike McDowell's claim for manifestation of meanings, the ordinary speaker cannot be said to grasp the configuration of c-command as a result of knowing that someone is asserting a sentence exhibiting that configuration. No such grounds for ordinary intentional attribution exist in the case of the ordinary interpreter.

Does this mean the theorist must be non-modest about these notions; ie. that he must introduce them by means of a theory that explains them rather than supposing that they were familiar everyday notions for ordinary speakers? McDowell certainly wants to resist this if he can. There is just one avenue left open to him: his account of the epistemology of understanding as a form of perceptual knowledge. With this direct perceptual account of understanding we can avoid the detour through the interpretation theory and ask what those perceptual states involve.

Do we bring syntax to what we read and hear? Perhaps modesty in this context means that we have to impose the syntax of our own language on the languages of those we interpret, just as Davidson thinks we read quantificational logic into the meanings of their sentences. Perhaps this is what David Wiggins means when he talks of seeing meaning through syntax. Speaking about the misleading properties of surface structure, he says:

Surely the surface can only mislead those who misperceive (or misread) it. How much you can see in a place depends on what you know. If we take this thought seriously, then we may even find ways to re-direct Davidson's good question 'What are these familiar words doing?' and train it upon what is actually visible in language (visible to a knowing eye, that is.) (Wiggins 1986, p303)

With all of this I entirely agree. But Wiggins uses this to suggest that deep structure need no longer be seen "as the province of an utterly mysterious sort of speaker's knowledge" What I fail to see is how these tantalising remarks about what can be seen on the surface by a properly equipped speaker can lead Wiggins to the conclusion that "we shall have no need for the rudimentary ideas of 'grammar' and 'surface structure' that sustain our present distinction between the surface and the depths" (Ibid, p304). On the contrary, we shall have need for the idea that the resultant insight into the surface of speech demands a fusion of all of the

Chapter Eight

notions that take place at different processing stages and which jointly and severally play a crucial part in sustaining the speaker's 'vision'.

Neither Wiggins nor McDowell give us a way of characterising what is perceived vis-a-vis syntax. But we know that McDowell has always insisted upon there being a distinctive phenomenology of understanding: there is something it is like to perceive another as uttering a speech act with such and such a force and such and such a meaning. We need to know whether these perceptions include hearing the sentence as structured. McDowell's answer to this question is very important, for unlike Dummett, he believes that the whole of a subject's experience must be conceptualised by the subject: there are no non-conceptual experiences. (Personal communication. Also see Strawson 1966, p272-3). And what a subject knows in this conceptual sense must be something that can be ascribed to him by an interpreter and so, a fortiori, it must be something that can be recorded in an interpretation theory. But we have already seen the hopelessness of any modest construal of the syntactic concepts we need to make use of at that level. So it looks as though McDowell owes us an account of what we know of syntax let alone what our knowledge of it consists in.

Certainly, interpretation theory cannot explain the concepts of syntax in terms of the ordinary notions drawn from mental life, so it looks as if those concepts are ripe for explanatory relegation. At one point McDowell acknowledges that to have the capacity whose input-output relation is described by the theory, one does not have to know any such theory, at least not in any ordinary sense of "know" (McDowell 1977, p147). This is his cue to look elsewhere, though because of his views of the mental he is forbidden from taking it. It is not an ordinary sense of "know" it is a special sense of know for which we need a 'special science', like individual computational psychology. We do not really 'know' the theory but we do have some way of encoding this information. The theory describes certainly crucially relevant states that participate in our ability to produce and comprehend meaningful speech. So instead of supposing that the theoretical representations characterise linguistic abilities fully displayed in behaviour, we should suppose that theoretical

Chapter Eight

representations characterise the states and mechanisms causally responsible for the behaviour that reveals linguistic competence.

It is not strictly correct to say that we grasp the concepts of syntax; at least not if grasp is taken to mean everyday concept mastery. And yet we want at least to ascribe grasp of syntactic structure to speakers of the language. How should we describe this? I think the trouble is that the notion of grasp is a place holder for various notions that show the varying grades of intentional involvement of the subject of experience. It may be that the notion of grasp and what we grasp spans the personal/ sub-personal divide. Certainly a theory of grasp of meaning will have recourse to notions like our grasp of structure and will refine the analysis of what constitutes grasp or possession of linguistic knowledge without heeding the personal/sub-personal division between kinds of cognitive state. But perhaps we need a detailed account of what this grasp depends on and consists in. This would require what linguists call a cline from informational reception to registration, recording, re-identification, to representation and all the way up to knowing. What, I think, the notion of linguistic grasp excludes is the concept of belief.

What all of this comes down to is just the problem we find with the notion of a speaker's knowledge of language. The complexity of that concept emerges when we try to satisfy two different requirements on it. The first is to make sure it encompasses all the things we take to be pertinent to a speaker's mastery of a given language, while at the same time trying to obey philosophical standards on what counts as a genuine state of knowledge. It seems to me that we should hold onto the idea that there is a large and complex property of knowing a language which may span different senses of 'know', and whose precise character it is the task of the philosopher of cognitive science to determine. This is preferable to simply missing the target of what we are trying to describe by insisting on a priori stipulations on what is to count as knowledge in the case of knowledge of language. Problems with the cherished analytical concept arise whenever we work on the notion of knowledge of meaning, rules, concepts or contents. Just to take one example, if grasp of the concept of an F is thought to involve knowledge of what it is for something to be an F, that knowledge had better not be given in purely propositional terms,

Chapter Eight

involving either the concept itself - for that is tightly circular - or those in terms of which we can give an analytic definition of the concept F - for this leads to a regress. This example alone should convince us that we need some more sophisticated tools. For this we must turn to full-blooded cognitive theories.

Lastly, there is the cognitivist or explanatory theorist's approach to syntax. This should be 'visible' already, but let me just draw one or two things together. Chomsky tells us that we should take the relational analysis of 'Jones know a language L' to be about the relation between a speaker and his I-language. This is a complex function that enables the speaker to interpret arbitrary expressions of the language of his community; the sort of function that McDowell hoped to characterise by the interpreter's theory of the information processing capacity. Chomsky tells us:

The I-language is intensional in that it is a specific characterisation of a function, considered in intension; it is internalised in that it is an element of Jones' mind, namely an element abstracted under the relational analysis [ie. Jones knows L] of the steady state of the language faculty. (Chomsky 1987, p181, my parentheses)

The steady state is attained from the initial state of the language after a period of maturation which involves learning vocabulary and setting parameters. The theoretical linguist's account of this already involves reference to internalised structures:

Taking a grammar as the theory of the I-language, the statements of a grammar as statements of the theory of mind about the I-language, hence statements about structures of the brain formulated as a certain level of abstraction from mechanisms.Statements about I-language or the statements that Jones knows I-language gE are true or false much in the way that statements about the chemical structure of benzene, or about the valence of oxygen,.....are true or false. (Ibid.)

Chomsky has no doubt about the linking principles for the theory of syntax; they relate the theoretical notions to cognitive properties of individual speakers. So he is committed to giving an account of both what we know and what our having that knowledge consists in. In my sense,

he is providing part of a theory of our grasp of what we know when we know how to speak a language.

Some will want to give a Marr level 1 account of the objects of knowledge which is also an account of our grasp of what is known. But at what level of the Marr hierarchy should we expect to find an account of our grasp of sense? Surely a theory of the computed function-in-extension cannot provide the right level of description of our knowledge of it. And yet McDowell, for example, would see himself as giving a level 1 theory of sense. It is not clear how this can be a theory of the actual knowledge of native speakers. It gives us no idea of how speakers possess the knowledge of the formal structures in the semantic theory. To do this it would also have to be a substantive theory of grasp. A substantive theory of grasp is more constrained than a formal theory of what is purportedly grasped by us. However, we should not straightforwardly assume that a theory of grasp is a level 2 psychological theory of the actual algorithms by which we arrive at an understanding of sentences. The case is clearer in syntax where a distinction is made between the grammar and the parser. People who may be said to have the same internalised grammars may still show differences at the psychological level of their individual algorithms for processing sentence structure. One may parse sentence structure from left-to-right, while another works from the middle out. Someone maybe following a top-down parsing strategy where one is looking for certain configurations and trying to match parts of the sentence against particular structures described by the grammar, while another may be using a bottom-up parser, resolving larger and larger constituents of the phrase structure described by the same grammar. Other issues in parser design involve whether we process in parallel, by trying all options at once when confronted with alternatives, or process in serial with backtracking when an option fails. All of these options suggest that while different parsing strategies are modelled by different psychological algorithms but they can all apply to the same grammar. The grammar does not fix the parser, hence the need for a grammar/parser distinction, and the need for a distinction between the theoretical level at which grammars are specified and the level at which we give psychological

Chapter Eight

theories of the algorithms for parsing. This in turn argues for the need to locate these notions at different levels in the speaker's cognitive system.

A theory of our grasp of syntactic structure, if it does not reside at level 1, and it needs to unite different algorithms at level 2, would have to be located at some inter-level. This is what Christopher Peacocke has supplied in putting forward his proposal for a Marr Level 1.5. (Peacocke 1986b) This can be thought of as an equivalence class of algorithms. Level 2 algorithms are equivalent at level 1.5 if they all draw upon the same information represented by a level 1.5 theory. At the same time, there can be different level 1.5 realisations of a level 1 theory; for example, the finitely axiomatised and the listiform (theorems as axioms) versions of a semantic theory for a finite language. Level 1.5 is the level at which to locate theories of grasp. Anachronistically, we can say that this is what Frege was after in his enigmatic remarks about a theory of our grasping being neither wholly psychological nor logical. Peacocke's level 1.5 seems to do the trick.

We can now use this notion to describe the first of two accounts of what it would mean to say that a grammar *G* is psychologically real. Peacocke puts it like this:

Take any statement *q* of grammar which is derivable in *G* from rules *R*₁,...,*R*_{*n*}; then not merely is it true that *q*, but the fact that *q* holds for the subject's language has a corresponding explanation at level 1.5 by some algorithm or mechanism in the subject, an algorithm or mechanism which draws upon the information that *p*₁, upon the information that *p*₂,...and on the information that *p*_{*n*}. (Peacocke 1989, p115)

This will do for the grammar of individuals but when linguists propose principles of grammar they are offering substantive universal claims about the mind. Peacocke has thought of this too:

...for a proposed principle of grammar to be psychologically real is for it to give a specifically linguistic content drawn upon, or a specifically linguistic transition-type used, in the individual's acquisition of a psychologically real grammar for a particular language (Ibid, p125)

Chapter Eight

This provides a constitutive account of what it is for a universal grammar to be psychologically real. Only linguistic states of subjects with these linguistic contents can play a part in claims for the psychological reality of principles of UG.

Peacocke's criteria show how specifiable grammars can enter psychological explanations. But before they can be empirically tested, what is urgently needed is an account of what it is for an algorithm to "draw upon the information given" for that notion has to be resolved before we get a definition of the equivalence classes of empirically testable level 2 algorithms. The right notion lies somewhere between the two extremes of having to have the principles, rules and structures explicitly represented in the system at one end, and it merely being the case that these rules describe the inputs and outputs of the system (ie. merely that they are true of them). Neither of these is quite right, for the notion of the information drawn upon cannot always be fully represented, some of the rules will be merely part of the computational architecture of the system, transitions it is the system's function to carry out which are not explicitly represented. Whereas to say that some system is drawing upon certain information in its functioning just because that information is true of it is too permissive to count as a condition for 'mental representation'. For as James Higginbotham points out (1986): an algorithm that computes $x^2 - 1$ on input x could be said in this weak sense to draw upon the information that $x^2 - 1 = (x - 1)(x + 1)$, since it relies upon this being the case for its success, without the algorithm representing this information, or transforming the problem into that of multiplying $(x - 1)$ by $(x + 1)$. I suggest that what we need here is a distinction between an algorithm 'relying upon information being the case' and it 'relying upon that information for its successful operation'. We need a substantive account of that second notion which does not insist on all the information being explicitly represented in the system. More needs to be said but Peacocke has introduced an important level of computational explanation and a promising explanatory notion. This is the notion we need if the internal organisation of a theory of syntax and the deductive organisation of a truth theory are to give rise to empirical hypotheses about the psychological explanation of speakers' competence.

Chapter Eight

The other account of psychological reality is the application of Davies' notion of causal systematicity to the notion of our knowledge of linguistic rules. Davies defines the latter in terms of the former and thus provides a constitutive account of **tacit knowledge**. The account would go like this. When there is a causal systematicity relative to patterns brought to light by high-level syntactic-theoretical description of the input-output states of a processing system then that system has tacit knowledge of the generalised rule that describes the pattern. The inputs can be mere emissions of noise and the output states can be recognitions of well-formed sentences (grammaticality judgements). So, for example, when we come across sentences with the form of our example wh-sentence:

(S) Who did Bill see?

as inputs, and we always get as output states describable as having the same form as knowing that "who" is the displaced direct object of "see", then, if an individual's linguistic processing system is causally systematic with respect to the rule of wh-movement on objects, then the system embodies tacit knowledge of the rule of wh-movement. And as we know this does not require that wh-movement be explicitly represented in the system; it only requires that there be states of the individual which encode the structures the rule describes. Perhaps the claim that certain internal states of competent speakers have a syntax is all that is meant by the Language of Thought Hypothesis. As Fodor puts it, programs can be inexplicit, but this does not show that everything else can be too. A parser which embodies the rule wh-movement and which transforms D-structures to S-structures for sentences like (S):

...can't compute without representing these structures because representations of structures are what it computes on.The moral: programs needn't be explicitly represented, but computational domains (roughly, data structures) have got to be. So its computational domains, not programs, that make the argument for the language of thought. (Fodor 1987, p75-6)

This would mean that there has to be a system of internal structures in the agent, structures which belong to a cognitive level of description of

Chapter Eight

the agent, which mediate linguistic inputs and outputs, which are not conceptualised by the subject, and which somehow come to acquire content, perhaps by means of certain complex relations between the subject and his environment and community. In such a system, it may be, as Davies says, that a component of the processor is responsible for our standing knowledge of a rule such as *wh*-movement. (Davies forthcoming)

Whichever notion of mental representation we chose to adopt, these psychologically relevant accounts have the virtue of incorporating claims about the internal cognitive organisation of the language user into claims about the grammar the individual possesses. In addition, they show claims about the psychological reality (or tacit knowledge) of grammars to have empirical content. So if it is one's aim to construct a grammar for the speakers' languages that comes closest to capturing the actual syntactic properties of the language they actually speak, then one is embarking on a psychologically relevant project. One is making empirical psychological claims. Patterns of breakdown, specific lesioning of states, and so forth could show how competence breaks down with respect to particular rules, or particular words, thus showing that there is a common empirical explanation underlying our understanding of a particular word or certain constructions. To avoid irrelevancy and false claims about speakers' knowledge of words and knowledge of semantic and syntactic structure one should pay heed to the available empirical theories of languages and cognition for those speakers.

Although I have suggested that grasp of structure may span the personal/sub-personal divide, involving both what is consciously, or phenomenologically, accessible and certain processing states we can be in, we have to respect the distinction for philosophical purposes. Davies' notion of tacit knowledge picks out states of a subject that differ from ordinary propositional attitude states in several important respects. Firstly, the contents of states of tacit knowledge are not conceptualised by the subject in those states. Linguistic rules may provide important shaping constraints on linguistically expressible thoughts, they themselves are not contents of our thoughts. Another important difference is that whereas propositional attitudes are inferentially

Chapter Eight

integrated, tacit knowledge of the principles of grammar have their own proprietary domains of application: the syntactic structures of sentences and expressions. Consciousness has often been used to differentiate states of tacit knowledge and states of belief, but consciousness is itself too imprecise an instrument to help us much with our characterisation. There are at least two quite different notions of the unconscious, one involving an essential subjectivity, the other failing to incorporate any conception of the subject's point of view. These must be kept distinct. It is unclear whether the distinction between conceptual and non-conceptual information will do service here since it is unclear whether we should describe the first notion of the unconscious in terms of propositional attitudes at all. These unconscious states seem to be involved in actions; they can share features of both belief and desire and yet lack the logical structure of either. (cf. where wishing something is to imagine it is true.) This suggests that there are many different kinds of cognitive state to consider, some of which are more, and some of which are less, complex than beliefs and desires, some of which that are both more and less complex than beliefs and desires along different axes. Among both the Freudian and the information-processing notions of an unconscious state there will be some that show the structure of belief but remain inferentially unintegrated, giving rise to no other beliefs and insulated from the effects of changing our beliefs about the information they contain. A proper understanding of our cognitive lives would demand a large scale study of this variety of cognitive states and the differences between them. What is important about this notion of cognition is that it includes much more in the domain of the mental than the Davidsonian philosopher of mind's restriction to the propositional attitudes. We have seen the explanatory pressure towards the introduction of the broad domain of cognition, and although I have offered no detailed philosophical psychology here, there are still one of two morals to be drawn.

One of the important things to remember when working with a personal/sub-personal distinction is that sub-personal properties must not be ascribed to the subject, nor should personal level properties to ascribed

Chapter Eight

to the cognitive subsystems. McDowell confuses these quite frequently in "The Sense and Reference of Proper Names" Consider the following:

In order to acquire an information-processing capacity with the right input-output relation, it would suffice to know the theory; then one could move from input information to output information, on any of the relevant occasions by explicit deduction. (1977, p147)

It is not obvious that this is true. One may have reflective knowledge of a theory of some computational task without being able to perform that task. Presumably, a blind person could read David Marr's account of visual processing in Braille without thereby gaining the capacity to see. We are not required to bring to bear a theory, although somehow unconsciously in understanding a visual scene; rather, we are required to possess the information wrought by certain processes which apply internally represented information to the raw input. To have structured vision of a scene we don't have to know a theory of the states of the visual processing system, we have to be in them.

Similarly, if the theory of language has to incorporate the details of our cognitive grasp of structure, it might be far from clear that knowledge of the theory conferred the capacity upon us. Though it is clear that without having the capacity already one could not understand the theory. But that is not to say that knowing the theory is a way of acquiring the capacity. The upshot is that understanding the theory entails having the capacity to understand the spoken and written language in which the theory is expressed. But it is far from clear that knowledge of the theory suffices for, in the sense of conferring upon the knowing subject, possession of that capacity. Instead we should say that anyone who understood the concepts used in the theory would already have rather than could gain the capacity. Once again, we don't need a theory of the capacity, we need the capacity.

Another example is provided by McDowell's inability to see what is intended in Dummett's insistence that there be a way of meaning something which is not a matter of interpretation. McDowell thinks that the only way Dummett can resist the interpreter's position is to suppose that speakers have knowledge of the psychological mechanisms by which

Chapter Eight

they arrive at an understanding of an utterance. But Dummett is supposing that people might possess, say recognitional capacities, not that they should know how they work. Failure to see this leads McDowell to criticize the more explanatory accounts as if they were erroneously providing theoretical details that a speaker/thinker had ordinarily to know to use language and to make sense of other speaker/thinkers. But this is ill-conceived. Speakers meaning what they do, or thinking what they think, is not a matter of the speaker/thinkers knowing the facts of how they do this, but it is a matter of there being certain facts true of them, or their being in certain states. These are facts which can be objects of theoretical knowledge for the explanatory theorist, but not for the ordinary subject. It is just this point that McDowell misunderstands, and which leads him to make the following misconceived criticism of attempts to construct richer, cognitivist conceptions of what speaker's knowledge of meaning consists in:

One can have the ability to tell that a seen object is the bearer of a familiar name without having the slightest idea how one recognises it. The presumed mechanism of recognition might be neural machinery - its operations quite unknown to its possessor. (Ibid. p147)

The answer to this supposed criticism, whether we are considering "neural" or cognitive mechanisms is simply, 'Of course!'

We have reached the conclusion that as far as our grasp of syntactic structure is concerned an account of what we grasp must go beyond the familiar concepts ordinary speakers can be presupposed to have just in virtue of speaking a language. This has the consequence that a theory of our knowledge of syntax must be non-modest, and we saw that a full-blooded theory of syntactic knowledge would involve more than a theory of patterns in our behaviour. A full-blooded theory of syntactic knowledge is a substantive theory of the character of certain task-specific processes and structures; it is a theory of speaker's cognitive states.

Conclusions and Reflections

I have now completed a part of what I take to be a much larger project: to determine the nature of the collaboration between philosopher and cognitive scientist in the study of language. The task here has been to discover the extent to which empirical considerations bear on the concerns of the theorist of meaning.

I have argued that any theory of language that serves as a theory of speakers' knowledge of language must include a psychological theory of the linguistic abilities of speakers. In this work, I have concentrated on the motivation for such an account in the case of syntax. It was argued that any descriptively adequate theory of language must include a grammar that postulates 'empty constituents' and syntactic configurations belonging to the underlying structure of sentences. These are phonetically unrealised elements belonging to structures in the syntax that carry information about grammatical relations within a sentence and they are motivated in the theory by the gap between what an utterance, or inscription, itself makes available, and what is necessary for determining a unique semantic content for the sentence heard, or read. However, they are not observable aspects of the surface of linguistic behaviour, nor are they known explicitly by speakers of the language. I then gave an argument from best explanation for ascribing speakers tacit knowledge of these posited syntactic structures, offering corroborating evidence for the psychological reality of syntactic forms with traces which showed their causal efficacy and perceptual impact on the production and reception of speech. I concluded that such syntactic elements are part of what is involved in understanding *which* sentence has been uttered on a given occasion, and in deciding what meaning it can be given.

The philosophical points to be drawn from this can be brought into focus by returning to Wright's remark that it is a platitude to say that:

if someone understands the vocabulary and syntax of a significant sentence, they understand the (type) sentence.

but that:

Conclusions and Reflections

Platitude is left behind when the antecedent of this conditional is taken to describe an ulterior state of information which *enables* a subject to understand the sentence. (1987, p208)

I have taken an explanatory approach to matters of understanding and argued that we need to develop the resources of a theory of cognition both to discover the facts about a speaker's internally represented grammar and to determine the role they play in explaining someone's linguistic mastery. The case I have made out, and the arguments advanced, are principally concerned with the *motivations* for the explanatory approach, and I have been able to offer only the briefest of sketches here of the foundations of a satisfying model of psychological explanation.

The arguments motivating the explanatory approach proceeded by highlighting the inadequacy of alternative approaches; in particular, pointing out that without explanatory details of actual speakers' linguistic competence they were unable to recover well-attested facts about actual languages. The issue turns on the question Wright asks when faced with talk of 'ulterior states'; is there any project which avoids recourse to the notion of speakers' tacit knowledge for which a compositional theory of their language is required? I could see just two (very different) cases to consider: the interpretive stance, and the descriptivist stance. I assessed each approach with respect to the nature of actual speaker's knowledge. The former position was founded upon a particular a priori conception of the nature of minds and their place in the world, which I referred to as attributionism. According to this view, an account of what makes knowledge of minds possible tells us all there is to know about the mind; and so by reflection on the everyday nature of psychological ascriptions we arrive at a constitutive thesis about what minds are. This renders the mind an attributed property in the eye of the interpreter. This leads the interpreter to say that the beliefs and meaning he attributes to other people are products of his attempt to rationalise their behaviour. This theorist has an answer to Wright's question which shows that there is no need on such an account to motivate a compositional theory of meaning by ascribing tacit knowledge of semantic theory to speakers. Instead the semantic structure in a language is imposed on it by the exigencies of conducting radical interpretation in a systematic way. But to accept this

Conclusions and Reflections

solution one would have to embrace its attributionist conception of the mental. I gave a full account of what such a conception involves.

Minds, as Davidson conceives them, turned out to be causally inert; and thus while Davidson may be a realist about propositional attitudes, he is not, like Fodor, a causal realist about such states. However, the realism about beliefs did not extend to concepts; Davidson turned out to be an instrumentalist here. Like the words which express them, concepts are just posits internal to the theory which have no empirical significance. Furthermore, the only real claim to conceptual structure is based on the structure of language, and this is the externally attributed property due to the project of radical interpretation. Finally, there is extensive indeterminacy in the mental due to the availability of equally adequate alternative schemes. We can ascribe beliefs but it is unclear precisely which beliefs we have. This position, which claims to adhere to commonsense more closely than any envisaged scientific psychology in the end proves highly revisionary of our ordinary views of the mental. Caveat emptor.

I showed that although there was no way to account for the success of radical interpretation and a speaker's susceptibility to the interpreter's descriptions of his language and mind, no such explanation was necessary on the attributionist picture. My worry was that this essentially third-personal account of the mental could not accommodate the notion of actual speakers' knowledge of language satisfactorily. And while it could be said that this was unnecessary since the interpretation theory attributes such knowledge, there was no satisfactory account of the interpreter's understanding of the language of the theory. To explain this by appeal to the possibility of being interpreted from the third-personal stance is not to explain the actual capacities of the speaker/interpreter.

In the end, my objection to such an approach was not that it left it unexplained and surprising that physical creatures should satisfy these high-level intentional descriptions; quite the contrary, I argued that to make these incontestable claims about our intentional activities certain empirical generalisations have to be true of us already, and it is a mistake to suppose that these can be left out of the account of what we do and say. In this respect, it is syntax which constitutes the most serious problem for

Conclusions and Reflections

the Davidsonian. For if there is no notion of a speaker's knowledge of his own language, we must suppose that syntactic information about a language, like semantic structure, emerges from the project of radical interpretation. I pointed out the implausibility of supposing that it could be the radical interpreter who engaged in detailed syntactic theorising about underlying levels of structure. For he is supposed to be engaged in a project that begins once he has made an initial identification of the sentences held true by the speakers he interprets. We are given no very clear indication of how he does this, beyond his being equipped to recognise certain linguistic forms already. However, there is enough evidence to show that if he had to extract the sentences of the language of the studied community, he would have to frame highly sophisticated hypotheses about the phonological, morphological and syntactic structure of that language; and for this he would need a medium in which to record his conjectures. This merely presupposed an antecedent grasp of (a good deal of) syntax, and fails to tell us anything about the ordinary speaker or the language learner's knowledge of language. Worse still, the interpreter is limited by two important restrictions: he can only observe the surface of another's behaviour, his assents and dissents; and he can only use materials minimally richer than those available to the ordinary speaker. It is to be doubted for the reasons offered above that any such account could succeed without embarking on an empirical inquiry that would oblige him to provide an explanatorily adequate grammar of a speaker's language; ie. one that postulated empty categories and called upon them in describing the psychological operations causally responsible for linguistic competence. This caused trouble for the radical interpreter in two ways. Firstly, as an attributionist about minds and meanings, he believes them to be properties externally attributed to speakers on the basis of their behaviour; this makes the postulation of internal psychological states deeply problematic. Secondly, even if the radical interpreter were to pick out the causally relevant states of speakers by an external attributed syntactic description, these ascriptions would not be specifiable by, or answerable to, the a priori constraints of rationality; on the other hand, they are not obviously descriptions of merely physical mechanisms. So the trouble for the radical interpreter is that he has no motivation for

Conclusions and Reflections

providing a descriptively adequate syntax; and without one, he has offered us too little reason for thinking he can effect the transitions from the physical to the semantical level.

Against the attributionist interpreter I have argued that we need to recognise other levels of cognition, intermediate between physics and folk psychology and reducible to neither. This is necessary since our grasp of syntax is neither a matter of reason nor causes alone. Pace the radical interpreter, to make initial identifications of someone's sentences, if they are correct, is to have entered the mind of the speaker already. It is to identify states of linguistic understanding that presume upon the cognitive organisation required for speech.

The same challenge is faced by one who seeks to *describe* the actual facts about a speaker's understanding in a way that falls short of *explaining* linguistic competence. It is unclear what the descriptivist's conception of mind is. If it is an a priori conception like the attributionist's, the two positions collapse. If it is some other a priori conception then we have still to be told a good deal more about what such a position might look like. A fully worked out picture of Wittgenstein's conception of mental life might serve here; but I have not tried to examine that suggestion in this work. I suspect that the descriptivist's alternative to attributionism tries to establish facts about minds as being revealed in language, which it then seeks to describe without appeal to inner states or empirical underpinnings. For this theorist there must be a non-attributionist account of the facts which constitute a speaker's competence in his native language, it is then the task of the meaning theorist to describe these facts as faithfully as he can. Typically, these will be conceived as facts about the speaker's use of a language, and the meaning theorist will aim to provide "an extended description of what that use consists in" (Dummett 1973, p681). But to characterise the *knowledge* involved in language-mastery, the theorist's description of use must not merely amount to a summary of the unvarnished facts about a speaker's behaviour. The theory of meaning for a language must be a theoretical representation of its speakers' knowledge. But if speakers do not themselves know such a theory what does their knowledge consist in? As we saw, speakers are ascribed implicit knowledge of the theory in the form of a complex ability to use and

Conclusions and Reflections

understand the language. The meaning theorist then appeals to the structure of the theory to articulate that complex ability into its component parts. The descriptivist's claim to be describing the *actual* structure of a speaker's linguistic abilities, and thus the structure of *actual* knowledge, rests on the claim that the speaker can manifest these pieces of knowledge in his use of particular sentences and expressions. However, we saw reason to doubt that a full-blooded theorist who wanted to describe episodes of behaviour as the manifestation of grasp of structured meanings could both maintain his non-attributionist stance and avoid adverting to the psychological structure of these linguistic abilities. In a similar vein, I questioned Wright's entitlement to suppose that there is a set of structured abilities standing at one remove from the behaviour in which they are conjointly manifested. To justify this claim one must either be an attributionist theorising about the form and content of linguistic abilities, or else be prepared to give a substantive account of them on the model of our cognitive psychological abilities. Certainly, in the case of syntax much more must be said about the structure of our linguistic abilities if we are to substantiate claims for an internal structure to sentences, and if we are to be able to describe a speaker's behaviour as *manifesting* his grasp of the underlying structures. In the end, I concluded that the descriptivist is trying to occupy an unstable middle position. And since I have given reasons for thinking that the interpretative stance fails both to recover what speakers actually know about their language and ascribes to them theoretical knowledge they don't have, I have advocated the adoption of an explanatory approach to language mastery.

In the larger project there is much is left to do. So far we have only considered conceptual structure and syntactic structure; next one must turn to semantics. In the first two cases I argued that we can think of conceptual and syntactic structure as being in the head. But when we turn to meaning we cannot argue by analogy to a similar conclusion. There are just too many differences between meaning and grammar to allow us to do this. Unlike syntax, the referential properties of expressions relate them to the world. The notion of an expression's meaning has normative force for one who grasps it: he must keep faith with the meaning he attaches to it. All these differences between the cases concern the content of

Conclusions and Reflections

expressions and not their form. So I suggest that we should think of a semantic programme as involving two parts; one dealing with semantic structure, and the other concerned with lexical semantics. The first component of this programme could be treated on analogy with the other cases. For unless we were to recapitulate the systematic relations between meaning and grammar once again, outside the head, there will be a problem of how to describe the interface between grammar and semantic structure. Moreover, it would be quite wrong to ignore the close relations between these different notions of structure, and quite false to persist in believing in the autonomy of syntax claim. Contemporary syntactic theories have long since abandoned the idea of completing a syntactic description of a language independently of its semantics. We saw in Chomsky's GB that he makes use of Binding Principles which govern co-reference relations, and appeals to a level of Logical Form as a level that records a partial representation of meaning including information about the scope of quantifier phrases. These are genuinely semantic parts of the grammar: a fruitful area for investigation of the relations between syntax and semantics.

If semantic structures along with the other notions of structure are in the head, then the form of a theory of meaning would be (in part) an empirical hypothesis. For I have been keen to stress the empirical nature of these structures and the fact that they are susceptible to ordinary standards of scientific investigation. To confirm the correct form for a theory of meaning for a set of speakers one would need to know something about their internal cognitive organisation. In particular one would have to appeal to a causal explanatory structure of content-bearing states underlying their competence in the language; and in particular, one would have to find out whether there was a uniform explanation for all speakers who hear a sentence as containing certain singular terms and predicates or having a certain semantical structure. However, these considerations cannot explain the substantive issues about semantical content. For this we need another sort of account.

Beyond the issues of grasp I have been concerned with lie questions about the particular nature of contents grasped. Like Dummett, I would favour a molecularist construal of the theory of meaning. I have not

Conclusions and Reflections

argued for molecularism here, but if one were to do so, the argument would have to be based on Dummett's progressive acquisition thesis which "allows for the arrangement of sentences and expressions of the language in a partial ordering, according as the understanding of one expression is or is not depending on the prior understanding of another" (Dummett 1976, p79) This is a partial ordering with minimal elements, and since the linguistic minimal elements are recognition statements, which, I argued, must have components corresponding to the components of a recognitional capacity, we have to retreat to the theory of thought to account for the conceptual content expressed in such statements.

I have not argued explicitly for molecularism, but there is still an important conditional claim to be made here. If one wants to adopt the position of a semantic molecularist, then one must provide a non-attributionist account of what our knowledge of the meaning of individual sentences consists in which makes room for the internal structure of basic statements. Also, an account that establishes the psychological reality of individual sentence meaning will offer strong grounds to contest the Davidsonian holist's claim that the attributed notions of individual belief and sentence meaning have a merely theoretical status within interpretation theory. To contest this one must find a level of cognition at which to locate the particulate facts of sentence understanding. At the level of thought it will be possible to describe a network of recognitional capacities that make up our cognitive map of our environment. The recognitional capacities will be more than just nodes in a network. To engender conceptual contents, satisfying Evans' Generality Constraint, they will have to be treated as structured abilities involving the exercise of component abilities which can be exercised in different combinations.

A full account of the structure and content of recognition based thinking is required to arrive at the correct account of the contents of the base class for progressive acquisition of language since Dummett appeals to recognitional capacities as an account of our understanding of statements in the base class. And as he says, the theory of meaning "ought to be modelled on our actual practice in coming to recognise statements as true". (Dummett 1973, p591) To account for the actual practice it is

Conclusions and Reflections

necessary to give an account of the content of recognitional judgements. The faculty of recognition operates to deliver knowledge-states whose contents are re-describable in truth theory at the level of output theorems, so if we want to know whether a truth theory ascribes the correct truth conditions to statements - the truth conditions speakers actually grasp - then we must give some reason for pairing each knowledge-state with a theorem of a particular truth theory. As Wright acknowledges, what counts here is not the manifestations in behaviour so much as the manifestable abilities in which possession of that knowledge consists. But there seems no way to work out what contents those cognitive capacities generate if we do not examine the elements involved in the recognition-based thoughts they support.

The recognitional capacities have a pivotal role in linking the individual speaker's use of language to the world and to other speakers. Both of these links are vital if one is to grasp the content of items in a public language. They provide a model of how knowledge of meaning is housed within an individual in terms of his possessing a capacity to detect certain observable circumstances, so if we could relate these recognisable circumstances appropriately to the semantical details on the one hand and to the cognitional details on the other, we might effect a hook-up between the truth theory for his language and the psychological states of the individual language user. But this is only one strand in the notion of a recognitional capacity. The other strand has to do with the social aspect of language use. Facts about meaning depend on facts about the communal standards of use. But should we suppose that in his unthinking use of these recognitional capacities a speaker is somehow appealing to these public standards to regulate his own linguistic use? At first it might seem difficult to accommodate both strands. But both are a necessary part of providing an answer to the question: 'How does a speaker's possession of certain recognitional capacities make it possible for him to have a grasp of the meanings of words and sentences in a public language?' For this reason we shall have to consider both individual and social aspects of language mastery. Without the psychological antecedents a speaker could not develop the abilities to make recognition-based statements; but without the linguistic community in which he finds himself, he would

Conclusions and Reflections

not acquire those very recognitional abilities which equip him to speak and understand the public language. In the course of linguistic development we have our recognitional capacities honed to the categories of perception and salience relevant to the needs and interests of our linguistic community. The task of mastering one's native tongue is that of tuning one's innately specified grammar to the parameter settings for one's community and tuning the references of one's terms to the references they have in the public language. The latter is achieved by acquiring the correct recognitional and inferential abilities, and trading on the finer discriminations of experts. The study of grammar in use and knowledge of reference is the largest part of the study of language mastery. It is a study involving both the social aspects of language use and the individual psychologies of language users. It shows both what equips one to participate in a practice and what shaping influence the practice has on what is acquired. We can settle for nothing less than this if we seek to give a substantive account of understanding.

Wright has written that for the antirealist:

...there is to be in the notion of understanding as little theoretical residue as possible in excess of what is necessary for description of the essential skills of the language-master. (1987, p276)

Given what I think has to be included even to describe the essential skills of the language master, is this account compatible with antirealism? I have not argued for antirealism, but I see no way in which what is said so far rules it out, so I take myself to be offering a position consistent with antirealist's views. There is no hidden appeal to knowledge of undetectable states of affairs, and no way in which semantical realism is assumed as a background assumption. All I insist upon is that the realism - antirealism issue should not be begged without considerations of our actual linguistic abilities. Since I have argued that these have to be illuminated by the psychological structure within an individual and conceptual elements described in the theory of thought, it is important that the move to the level of thought should not beg questions on the realist - antirealist issue either. But it doesn't. In fact, it is only at the level

Conclusions and Reflections

of thought that one can see the differences in the way the realist and antirealist conceive the truth conditions for the basic statements.

At the level of language use the realist and antirealist seem to coincide, but the contents grasped by a speaker have to do with his conception of the recognisable state of affairs that make those statements true. It is here that the realist is obliged to say how the judging subject can recognise not only that such a state of affairs obtains but also what it would be for that state of affairs to exist undetectably; and the antirealist will reserve the right to cast doubt on any account of thought which confers grasp of this extra element on the thinker. Realists like Evans and Peacocke can only establish this extra ingredient of the judgement by illicit appeals to realist assumptions. So for a basic statement like "That block is cubic", Evans sees a fundamental level of thought lying behind the the antirealist's fundamental level of recognitional thinking. At this level the subject thinks of the block from no point of view (1982, p152) The antirealist is likely to find "the view from nowhere" unintelligible, but he can still provide his own account of a more objective mode of thinking about objects in the world by reference to "the view from anywhere". A different suggestion which attempts to respect these cognitive limitations is offered by Peacocke who gives a realist account of what it is to grasp the objective truth conditions of the the thought that the block is cubic at time t in terms of a spectrum of canonical commitments the thinker undertakes to accept that from any position from which one were to perceive the block at t in normal conditions, it would look cubic or look as a cubic object would look from that position. To judge the thought is to accept these conditions, and the acceptance conditions determine the objective truth conditions of the judgement that that block is square at t . (see Peacocke 1986a, pp11-27) However, Peacocke establishes the extra ingredient for realism by means of the universal quantification used in specifying the canonical commitments; for he is quantifying over an infinite number of positions from which one could survey the block. However, no such infinite generalisation is necessary since the only relevant positions for these supporting expectations in perception are the ones which are cognitively discriminably different from one another to the human subject. Again no case for realism has been made out at the level of thought for the grasp of

Conclusions and Reflections

a basic statement's truth conditions. So I conclude that the issue of whether to be realist or antirealist about the content has to be disputed at the level of thought; merely moving to this level is not to concede the point to the realist.

One can now see how the account of language mastery I am advocating departs from Dummettian antirealism: it includes elements from the theory of thought in a theory of language; and it makes essential appeal to psychological facts about language users. I see these modifications as causing no disturbance to the main thrust of Dummett's position in the theory of meaning and as a necessary step for him to take to relieve his difficulties over the recognitional capacities. Appeal to the theory of thought does not replace the theory of language. There is no reversal of priorities here. Even the psychological component in this account cannot provide for the significance of speech of its own; it is merely there to illuminate the conditions on being the user of a public language.

The recommendation of this position is that Dummett needs to acknowledge not only the epistemological component in the theory of meaning, but also the psychological component. To date, Dummett has been unwilling to do this, and for largely Fregean reasons. It has been a salutary part of Dummett's enormous contribution to the philosophy of language to remind us that:

...the general contention that Frege wanted to extrude everything epistemological from logic or from the theory of meaning is quite misconceived: he wanted to extrude everything psychological. (Dummett 1973, p240)

However, once we have divested ourselves of mental images, private sensations and the methodology of introspectionism, it is important to stress that the theory of meaning should not extrude everything cognitive psychological just because it should extrude everything psychologistic. Our grasp of the syntactic structures and categories of the underlying levels of syntactic representation are not objects for private study, for they are not items for personal inspection at all. They belong to the sub-personal domain and we have this information in virtue of certain cognitive states we are in. With the need to account for the individual, social and

Conclusions and Reflections

psychological aspects of language mastery, we need to locate our account against a background conception of mind that accommodates the first-personal, third-personal and sub-personal aspects of mental life. For this we need a mature cognitive science.

Bibliography

- Bilgrami, A. 1989 Realism without Internalism: A Critique of Searle On Intentionality, *The Journal of Philosophy*, vol. 86, pp 57- 72.
- Burge, T. 1989 Wherein is Language Social?, in George, A.(ed.), *Reflections on Chomsky*, Oxford: Blackwell, pp. 131-52.
- Chomsky, N. 1972 *Language and Mind*, New York: Harcourt Brace Jovanovich
- 1975 Knowledge of Language, in Gunderson, K. (ed.), *Language, Mind and Knowledge*, Morris: University of Minnesota Press, pp. 299-320.
- 1981 *Lectures on Government and Binding*, Dordrecht: Foris.
- 1986 *Knowledge of Language: Its Nature, Origin and Use*, New York: Praeger Publishers
- 1987 Replies to Alexander George and Michal Brody, *Mind and Language*, Vol 2. no.2
- 1989 [forthcoming] *Language and Mind*, Darwin Lecture, Cambridge.
- Churchland, P 1988 *Matter and Consciousness*, Cambridge Mass: MIT
- Clark, H. 1986 Is Philosophy of Language Empirical?, *Mind and Language*, vol 1, pp15-17.
- Davidson, D. 1980 *Essays on Actions and Events*, Oxford: Clarendon Press.
- 1982 Rational Animals, *Dialectica*, vol 36, pp 318-327
- 1984 *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.
- 1984b First Person Authority, *Dialectica*, vol 38 pp101-111
- 1985 Psychology as Philosophy, in Vermazen, B. and Hintikka, M. 1985 *Essays on Davidson*, Oxford: Clarendon Press.
- 1985b Incoherence and Irrationality, *Dialectica*, vol 39, pp345-354.
- 1986(a) A Nice Derangement of Epitaths, in LePore, E. and McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson.*, Oxford: Blackwell.
- 1986(b) The Coherence Theory of Truth and Knowledge, in LePore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.

Bibliography

- 1987 On Knowing One's Own Mind, *Proceedings and Addresses of the American Philosophical Association* March.
- Davies, M. 1981 *Meaning, Quantification, Necessity*, London: Routledge and Kegan Paul.
- 1986 Tacit Knowledge and the Structure of Thought and Language, in Travis, C (ed.) *Meaning and Interpretation*, pp. 127-58.
- 1989 Tacit Knowledge and Subdoxastic States, in George, A.(ed.), *Reflections on Chomsky*, Oxford: Blackwell, pp. 131-52.
- [forthcoming] Concepts, Connectionism and the Language of Thought, *Proceedings of the International Colloquium on Cognitive Science*, Donostia-San Sebastian, 3-6 May 1989.
- Dennett, D. 1987 *The Intentional Stance*, Cambridge Mass: MIT/Bradford.
- 1988 The Intentional Stance, *Behavioural and Brain Sciences* Vol. 11 pp. 495-546
- Dowty, D., et al 1981 *Introduction to Montague Semantics*, Dordrecht: Reidel
- Dummett, M 1973 *Frege: Philosophy of Language*, London: Duckworth
- 1975(a) What is a Theory of Meaning, in Guttenplan, S. (ed.) *Mind and Language*, Oxford: Clarendon Press, pp. 97-138.
- 1975(b) What do I know when I know a language?, *Universitas Regia Stockholmensis*
- 1976 What is a Theory of Meaning? (II), in Evans, G. and McDowell, J. (eds.) *Truth and Meaning*, Oxford: Clarendon Press, pp. 67-137.
- 1978 *Truth and Other Enigmas*, London: Duckworth
- 1981(a) *The Interpretation of Frege's Philosophy*, London: Duckworth
- 1981(b) Objections to Chomsky: a review of *Rules and Representations*, in *The London Review of Books*, Sept 3-16
- 1982 Realism, *Synthese*, Vol. 52 pp. 55-112
- 1983 Language and Truth, in Harris, R. (ed), *Approaches to Language*, Oxford: Pergamon Press.
- 1986(a) A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking, in LePore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.
- 1986(b) The Philosophy of Thought and the Philosophy of Language, in Vuillemin, J. (ed.) *Merites et limites des methodes*

Bibliography

- logique en philosophie, Paris: Vrin et Foundation Singer Polinac, pp141-155.
- Dretske, F.I., 1980 The Intentionality of Cognitive States, in French, P, Uehling, T. and Wettstein, H. (eds.). *Midwest Studies in Philosophy*, Vol 5.
- Evans, G. 1976 Semantic Structure and Logical Form, in Evans G. and McDowell, J. (eds), *Truth and Meaning: Essays in Semantics*, Oxford, Clarendon Press, pp 199-222.
- 1981 Semantic Theory and Tacit Knowledge, in Holtzman, S. and Leich, C. (eds.) *Wittgenstein: To Follow a Rule*, London: Routledge and Kegan Paul, pp 119-137.
- 1982 *The Varieties of Reference*, McDowell J. (ed). Oxford: Oxford University of Oxford Press.
- 1985 *Collected Papers*, Oxford: Oxford University Press, 1985.
- Field, H. 1972 Tarski's Theory of Truth, *Philosophy*, vol LXIX.
- Fodor, J.A. 1987 A Situated Grandmother? Some Remarks on Proposals by Barwise and Perry, in *Mind & Language*, Vol 2, pp 64-81.
- Fodor J.A. and Bever, T.G., 1965 The Psychological Reality of Linguistic Segments, *Journal of Verbal Learning and Verbal Behaviour*, Vol 4, pp 414-420.
- Follesdal, D. 1982 The Status of Rationality Assumptions in Interpretation and the Explanation of Action, *Dialectica*, Vol. 36, pp301-316.
- Frege, G. 1884 *The Foundation of Arithmetic: A Logico-mathematical Enquiry into the Concept of Number*, tr. Austin, J.L. Oxford: Blackwell, 1959.
- 1979 *Posthumous Writings*, Hermes, H., Kambartel, F. and Kaulbach, F. (eds.), tr. Long, P. and White, R., Oxford: Blackwell.
- Fricker, E. 1982/3 Semantic Structure and Speaker's Understanding, in *Proceedings of the Aristotelian Society*, vol. 82.
- Grandy, R. 1973 Reference, Meaning and Belief, *The Journal of Philosophy*, Vol 70
- Grice, H. P. 1957 Meaning, *Philosophical Review*, Vol LXVI, pp 377-388
- 1968 Utterer's Meaning, Sentence-Meaning and Word-Meaning, *Foundations of Language*, Vol 4, pp 225-242.
- 1969 Utterer's Meaning and Intentions, *Philosophical Review*, Vol LXXVII, pp 147-177.

Bibliography

- Hacking I. 1986, *The Paradox of Conversation*, in LePore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.
- Higginbotham, J. 1983 *The Psychological Reality of Grammar*, in *How Many Questions: Essays in Honour of Sidney Morgenbesser*
- 1986 *Peacocke on Explanation in Psychology*, in *Mind & Language*, vol 1, pp. 358-387
- Hornsby, J. Which Physical Events are Mental Events?, in *Proceedings of the Aristotelian Society*, vol. 81.
- 1985 *Physicalism, Events and Part-Whole Relations*, in LePore, E. and McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson.*, Oxford: Blackwell.
- Johnston, M, 1985 *Why Having a Mind Matters*, in LePore, E. and McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson.*, Oxford: Blackwell.
- Lewis, D. 1974 *Radical Interpretation*, *Synthese* Vol. 23, pp. 331-44
- 1975 *Languages and Language*, in Keith Gunderson, (ed.) *Minnesota Studies in the Philosophy of Science*, Voll VII
- Marr, D. 1982 *Vision*, San Francisco: Freeman
- McDowell, J. 1976 *Truth Conditions, Bivalence and Verificationism*, in Evans G. and McDowell, J. (eds.), *Truth and Meaning: Essays in Semantics*, Oxford, Clarendon Press, pp. 42-66.
- 1977 *The Sense and Reference of a Proper Name*, in *Mind*, vol. 86, pp. 159-185.
- 1978 *Physicalism and Primitive Denotations: Field on Tarski*, in *Erkenntnis*, 13, reprinted in Platts, M. (ed.) *Reference, Truth and Reality*, Routledge and Kegan Paul, 1980, pp. 111-130.
- 1981 *Anti-Realism and the Epistemology of Understanding*, in Parret, H. and Bouveresse, J. (eds.) *Meaning and Understanding*, pp. 225-48.
- 1985 *De Re Senses*, in Wright, C. (ed.), *Frege: Tradition and Influence*, Oxford: Blackwell.
- 1985 *Functionalism and Anomalous Monism*, in LePore, E. and McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson.*, Oxford: Blackwell.

Bibliography

- 1987 In Defence of Modesty, in Taylor, B. (ed.), *Michael Dummett: Contributions to Philosophy*, Dordrecht: Martinus Nijhoff.
- McGinn, C. 1977 Charity, Interpretation and Belief, *Journal of Philosophy*, Vol 74
- 1981 Modal Reality, in Healey, R. (ed), *Reduction, Time and Reality*, Cambridge: Cambridge University Press.
- McLaughlin, B.P. 1985 Anomalous Monism and the Irreducibility of the Mental, in LePore, E. and McLaughlin, B. (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson.*, Oxford: Blackwell.
- Peacocke, C. 1979, *Holistic Explanation*, Oxford: Oxford University Press
- 1986 (a) *Thoughts: An Essay on Content*, Oxford: Blackwell
- 1986 (b) Explanation in Computational Psychology: Language, Perception and Level 1.5, in *Mind & Language*, vol 1.
- 1989u Content and Norms in Explanation and the Natural World. Unpublished.
- [forthcoming] What are Concepts? in French, P, Uehling, T. and Wettstein, H. (eds.). *Midwest Studies in Philosophy Vol 14: Contemporary Perspectives in the Philosophy of Language II*.
- Quine, W.V.O. 1960 *Word and Object*, Cambridge, MA: MIT Press.
- 1972, Methodological Reflections on Current Linguistic Theory, in Davidson, D. and Harman, G. (eds.), *Semantics of Natural Language*, Reidel, pp. 442-454.
- Root, M. 1986 Davidson and Social Science, in LePore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.
- Smith, B.C. 1982 Reflections and Semantics in a Procedural Language, PhD MIT.
- Stich, S. 1978 Beliefs and Subdoxastic States, *Philosophy of Science*, vol. 45, pp. 499-518.
- 1983 *From Folk Psychology to Cognitive Science*, Cambridge MA: MIT Press.
- Strawson, P.F. 1966 *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*, London: Methuen & Co.
- 1976 On Understanding the Structure of One's Language, in Evans, G. and McDowell J. (eds), *Truth and Meaning: Essays in Semantics*, Oxford: Clarendon Press

Bibliography

- Tennant, N. 1987 *Anti-Realism and Logic*, Oxford: Oxford University Press.
- Vermazen, B. and Hintikka, M. 1985 *Essays on Davidson*, Oxford: Clarendon Press.
- Weir, A. 1985 Against Holism, *Philosophical Quarterly*, Vol. 35, pp. 225-44
- Wiggins, D. 1980 What Would be a Substantial Theory of Truth?, in van Straaten, Z. (ed.) *Philosophical Subjects: Essays Presented to P.F. Strawson*, Oxford: Oxford University Press, pp 189-221.
- 1986 Verbs and Adverbs and Some Other Modes of Grammatical Combination, in *Proceedings of the Aristotelian Society*, vol. 86.
- Wittgenstein, L. 1953 *Philosophical Investigations*, tr. Anscombe, G.E.M., Oxford: Blackwell.
- 1956, *Remarks on the Foundations of Mathematics*, Oxford: Blackwell.
- 1958, *The Blue and Brown Books*, Oxford: Blackwell.
- 1969, *On Certainty*, tr. Anscombe, G.E.M., Oxford: Blackwell.
- Wright, C. 1980 *Wittgenstein on the Foundations of Mathematics*, London: Duckworth
- 1981 Rule-Following, Objectivity and the Theory of Meaning, in Holtzman, S. and Leich, C. (eds.) *Wittgenstein: To Follow a Rule*, London: Routledge and Kegan Paul; pp 99-117
- 1983 *Frege's Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press.
- 1986(a) How Can the Theory of Meaning be a Philosophical Project, in *Mind and Language*, vol. 1, pp 31-43.
- 1987 Theories of Meaning and Speakers' Knowledge, in 1987 *Realism, Meaning and Truth*, London: Blackwell, pp 204-238
- 1987 *Realism, Meaning and Truth*, London: Blackwell